6. FORECASTING METHODS

TABLE OF CONTENTS PAGE

Business data comes in many forms and in many types. The focus of analysis in most of the other chapters is with marketing or opinion research data. This chapter deals with general business data analysis. This consists of transaction, sales, financial and economic data. This chapter is intended to be more technical in that it deals with underlying problems and issues of the procedures some of which are mentioned in other chapters. Many of the problems discussed are still unresolved and most of the research results discussed are personal and do not represent a synthesis of all available information.

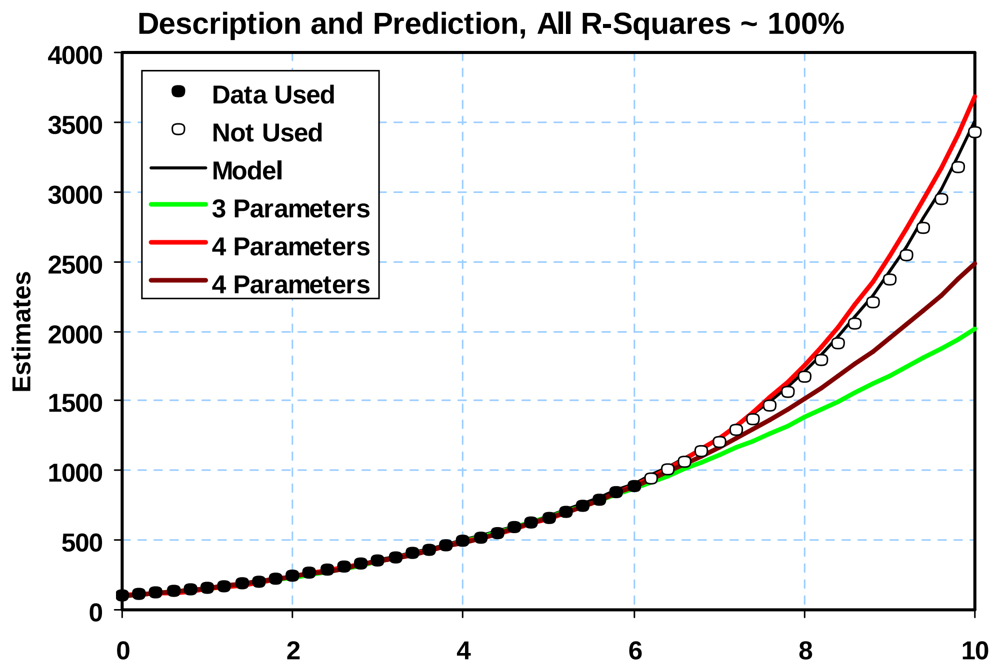
6.1 MODELS

All data analysis is based on underlying analytical “models.” These models are intended to provide a way to order the universe, provide insight and forecast potential future events. However, all models are not created equal. There are at least three different measures of model performance: (1) description of phenomena, (2) predicting future trends, and (3) insight into underlying mechanisms and deviations.

6.1.1 DESCRIPTION AND PREDICTION

The statistical quality of fit (R-Square) indicates the ability of a curve to describe data. This is the traditional measure for model selection. However, it is no assurance of the quality of a prediction. The belief that the better the descriptive fit, the better the forecast, is invalid. Below is an example of the problem[1]. Here three models were used to fit the data below a value of 6 and then extrapolated to over the rest of the domain. All model shown almost a 100% R-Square fit of the curves to the data over the region. However, the predicted values can be vastly different. In this case, one 4 parameter (polynomial) gave a very good forecast while another with almost identical R-Square value did not.

[1] The data was generated by a fifth order polynomial expression.

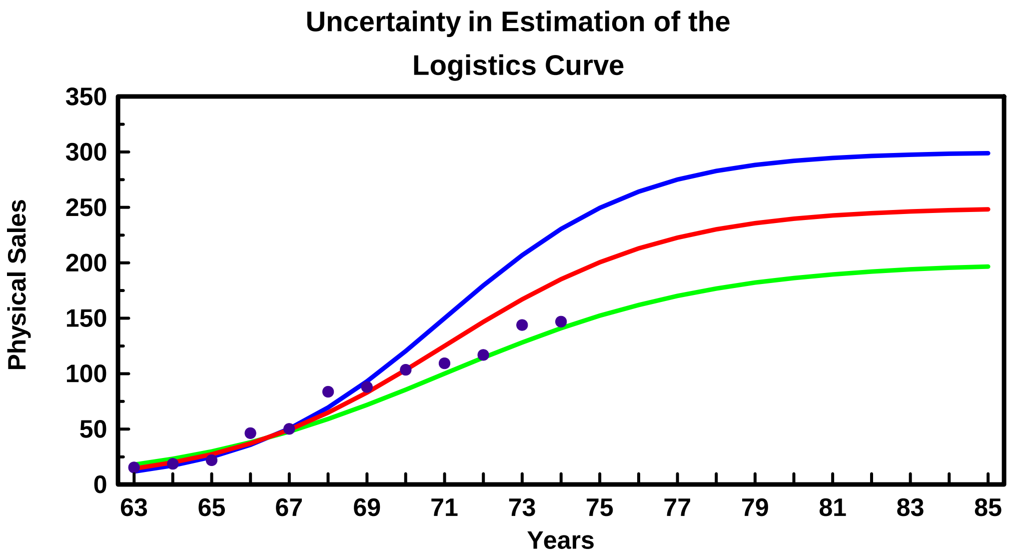


Particularly, for growth relationships with convex functions, R-Square measures can be very poor assessors of the ability of the curve to predict future values. This is not to say that the ability to describe data is not very useful, just that it may not be a good measure of predictability.

For some models, the difference between forecasting and describing the data can be critical. Logistics fit, needed for the Bass or Fishbein models, is very sensitive to the estimate of the upper boundary or plateau. Below is the case of three estimates of the growth curve based on a non-linear regression fit using increasing amounts of data. With more data, the curve drops. This would indicate a much higher estimate of the market with initial data. Logistics fits are notorious for this problem[2].

---

[2] This is particularly the problem with data below the inflection point of the curve (50% point). Unfortunately, data above that point is of less projective interest.



- **Forecast Sensitive to Estimation of Upper Bound**
- **R-Square Insensitive to Upper Bound**

It is therefore critical to test models for their predictive abilities rather than rely solely on descriptive goodness of fit.

6.1.2 FORMULATION OF MODELS

There are three general approaches to developing models based on: (1) mathematical forms, (2) underlying mechanisms, and (3) fitting phenomena. The use of simple mathematical forms is a standard approach based usually on the ease of computation. Linear, polynomial, logarithmic, power-law, exponential, or logistics[3] forms are used. Their selection is usually based on expected behavior and simplicity of the computation.

- **Mechanistic Models**

The scientific tradition lends itself to attempt to develop theory based on underlying mechanisms. This is a "reductionist" approach in that the process is believed to be cause by a simple mechanism for which we need only determine the underlying forces and parameters. Many of the mathematical forms, furthermore, lend themselves to mechanistic interpretation. Unfortunately, there are often any number of potential mechanisms that give rise to the same forms. Simplistic mechanisms rarely capture actual behavior and often prejudge the selection and development of models. Ideally, mechanistic models should only be a driver in the analysis if their underlying assumptions can be directly tested. Usually, this is not the case and the descriptive

[3] Transcendental functions, such as series of trigonometric functions (Fourier series) are also used but rarely for business data applications.

agreement between data and the resulting models are used as “validation” of the underlying model.

- **Phenomenological Models**

Most effective business models are based on describing data and should be considered phenomenological, in that they describe the phenomena rather than convey an understood mechanism. This approach is in the tradition of engineering rather than the scientific method. The objective is to describe what is happening and if feasible use forms that are consistent with underlying theory and mechanisms. However, the key is the proper and consistent description of phenomena.

6.1.3 FUNDAMENTALS VERSUS PRACTICE

It is important to recognize the distinction between fundamental relationships and those which are a result of specific practices. Fundamentals are invariant in principle though they manifest themselves through practices. The underlying relationships and definitions should remain constant over time and across applications. Practices, however, though highly consistent, are limited by time and use. New technologies, inevitably, effect practice but not the underlying fundamentals.

The major problem of phenomenological models is that they tend to confound fundamental behavior with operational practice. While mechanistic models, by their nature, focus on fundamental forces and issues, phenomenological models do not make this distinction. For this reason, it is critical that phenomenological expression should be tested over a broad range of time and conditions to verify generality.

6.1.4 INSIGHT

Effective modeling should provide insight into underlying behavior, which may not be apparent from the raw data. This is a “smoothing” effect where underlying forces are made more apparent with the reduction of “noise.”

Insight is derived from the parameter values and the conditions at which the models deviate from the data. Insight derived from parameters reflects the ability to interpret the model in terms of underlying forces. This is fairly straight forward with mechanistic models but is also capable with phenomenological models.

Insights derived from deviations are based on “piece-wise compliance” to the model. We assume that the model is valid over a range of the data. Noting deviation from the model identifies conditions where the underlying behavior has changed.

6.1.5 NUMBER OF ADJUSTABLE PARAMETERS

The number of adjustable parameters can effect the effective goodness-of-fit and the ability of the model to provide insight. The number of parameters may have no effect on the predictability of the model, as we have seen above. There are no general principles that more parameters produce more predictive models.

The larger the number of adjustable parameters, however, does result in a better description of data. An artifact of statistical fits is that as the number of parameters approaches the number of points available, the model approaches a perfect fit. The “Adjusted R-Square” measure corrects for some of this artificial improvement. However, the general problem of increased parameters monotonically improving data fit still remains. The statistical cut-off criterion for meaningful improvement is arbitrary and therefore, provides little guidance[4].

The major problem is using large numbers of adjustable parameters lies in the loss of insight. The larger the number of parameters the less meaningful is the individual values of the parameters and the less sensitivity to model deviations. As a general rule, we have found the fewer the number of parameters the better. This allows for simpler and better interpretation and easier data manipulation. We, typically use only two adjustable parameter models.

[4] Such criteria are used in Step-Wise regression procedures to determine the inclusion of additional parameters (variables).

6.2 PRODUCT LIFE CYCLE

The dynamics of sales are assumed to be dependent on the concept of the product life cycle with high early growth followed by a leveling off during maturity and finally a decline as other economically superior products substitute in appropriate applications. The traditional product life cycle concept contains a degree of inevitability. Eventually all technology are expected to pass through these stages. Unfortunately, some modeling procedures predict these stages to make the product life cycle concept a prediction of extinction. These models are not, however, reliable. While it can be expected that all products will mature and eventually decline, the timing is far from assured. Furthermore, the mechanisms for growth, maturity, and decline are usually quite different. Successful strategies differ markedly with product life cycle stages[5]. Each stage differs in dynamics and in return.

6.2.1 PRODUCT INTRODUCTION

Typically the product introduction stage is usually not included within the product life cycle. However, this initiation phase may be extremely long. Miss-introductions are not unusual. Products may fail in the first introduction or may take decades of development. The key problem in this phase is to identify when the product as actually become commercial. In some cases, commercialization takes place prior to release from laboratory production. It is only apparent from a historical perspective, in these cases, that commercialization had already taken place.

6.2.2 GROWTH

The well recognized stage in the product life cycle is growth. This phase may last as little as a couple of years to extend to a century as in the case of telephone long lines. We have found that this phase can be successfully modeled allowing for quality sales forecasts. This is discussed in detail below. However, it should be noted that the early stages of growth in the product life cycle is notoriously not profitable. New capacity investment requirements often outpace earnings making the business a continuous sink of capital. These “high successful” venture often do not become a net source of cash until the second decade of commercialization (average of 14 years). However, these businesses do become major sources of income for their firms later in growth and through maturity.

6.2.3 MATURATION

As some point most business sales tend to level off at least to the level of the growth in the economy. This is the maturity of the product in its life cycle. The key strategic problem with maturity is identifying when it has taken place. Mature business are more difficult to manage and rewards tend to be more limited. As such, management tends to

[5] Michael E. Porter, **Competitive Strategy, Techniques for Analyzing Industries and Competitors**, Simon & Schuster, Inc. (1980)

resist the concept that their business has matured and as such tend to try to manage them as if they were still in growth. However, mature business can be highly profitable once effective cost constraints are imposed. The causes of maturity include:

6.2.3.1 Product Saturation

The apparent and most often describe cause of maturation is the "saturation of the market." This refers to both the lack of new markets and applications and the inability to expand those existing. The key issue is what is behind the lack of continued growth. In general, it is due to the restriction in the increasing value of the product. New markets come into being when a significant value-added can be identified with the use of the product.

6.2.3.2 Sustained Value

Value is obtained by two mechanisms: (1) increase ability to effectively use the product, and (2) reduced prices. The first of these becomes a question of resources. During the growth phase, resources are typically available to pursue applications, modify products and develop relationships with customers. These activities tend to increase the value of the products in the market and makes and increasing they range of applications. In maturity, however, these resources are greatly reduced. A key question then is as much whether we wish to "mature" the business as it is whether the market has matured.

6.2.3.3 Cost Constraints

The inability to continuously reduce price or cost to the user is a key driver to product maturation. Between economies of scale and of experience, the costs of manufacture and to some extent of running the business should decline. In many cases input cost limit the cost reduction. This reduces the range of applications and tends to mature the business. However, it should be noted that the cost issue is on the use basis not on the specific production. Reducing product density and thickness can result in reduced cost products that may be highly acceptable for new applications.

6.2.4 DECLINE

The final phase is decline when product volume actually goes down. This is usually associated with further reduction in competition. While maturity may see some degree of substitution of products, decline always is associated with that substitution. This may be either in-type or functional in nature. In should be noted that decline can be highly profitable. During decline prices can rise. In addition, older, less productive manufacturing facilities are retired resulting in a net reduction in costs. This double effect pushes profit margins high even though volume continues to decline. This coupled with usually highly depreciated assets can produce a highly profitable situation.

6.3 FORECASTING TECHNIQUES

Forecasting, and in particular sales forecasting, are standard planning procedures in most businesses. While it is a normal activity in most annual planning and budgeting procedures, there are multiple reasons for its generation. The reasons behind its generation, effects both their required accuracy and biases. As part of the budgeting process, its major objective is to provide views of the resources and requirements that will be needed. It acts here as a basis for resource allocation. For marketing resources, short term corrections are feasible and; therefore, accuracy may not be critical.

Forecasts are also used as a manufacturing scheduling tool. In the case of long production times and periodic sales, these forecasts can be critical and often determine the inventory required. In these cases, there may be either an over or under estimate bias within the forecasting system. This is to assure either adequate supply or minimum inventory depending on the nature of the business.

Forecasts are also used to set sales objectives and are linked to sales and management compensation. Here again different biases may come into play. Finally, forecasts act as a measure of due diligence in the planning process. As such very conservative forecasts may be deemed unacceptable for an aggressive development plan such as with new product introduction. Under this condition, forecasts will tend to be biased on the high side. In all these cases bias is introduced by the process. It is, therefore, critical to first determine the expected uses for the forecast before undertaking the development of methodologies.

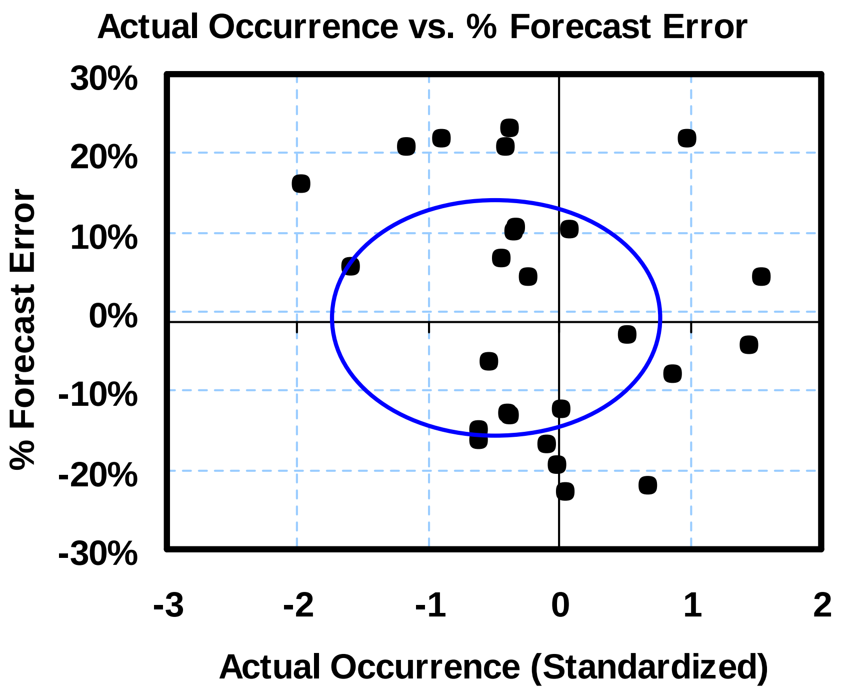
6.3.1 METHODS OF FORECASTING

There are two basic methods for sales forecasting: (1) “bottom-up” based on the opinions of field, expert, and management personnel as well as from customers as to what they expect to happen and (2) “top-down” approaches based on modeling historical data. Because of overlap in responsibilities and multiple sources, bottom-up data usually require adjustments[6]. Most practical procedures use elements of both methods, collecting subjective estimates from the field and comparing them with results from historical data and models.

6.3.2 FORECAST VALIDATION

Forecast validation consists of comparing forecasts with actual results. Typically there is significant variation. The key problems are: (1) the extent of the variation, (2) general bias, and (3) specific bias. A comparison of forecast versus actual is shown on the following graph designed to analyze for these problems.

[6] Informal adjustments are often used based on analysts judgement of consistency. With formal industry analysis, Delphi studies can be used to merge opinions.



This graph shows forecasts from several businesses over a number of years. Different years or businesses can be highlighted by using colors. The varying size of the businesses has been standardized. Deviation between actual and forecast is given as a percentage. The blue zone captures the variation in the data[7]. The wider this zone is from top to bottom characterizes the expected variation in the forecasts. The distance between the center of this zone from the origin of the plot characterizes the general bias in the results in this case it is extremely small, less than 1%. Distortion from a straight oval indicates the degree of special bias such as greater error with the larger businesses or business changes.

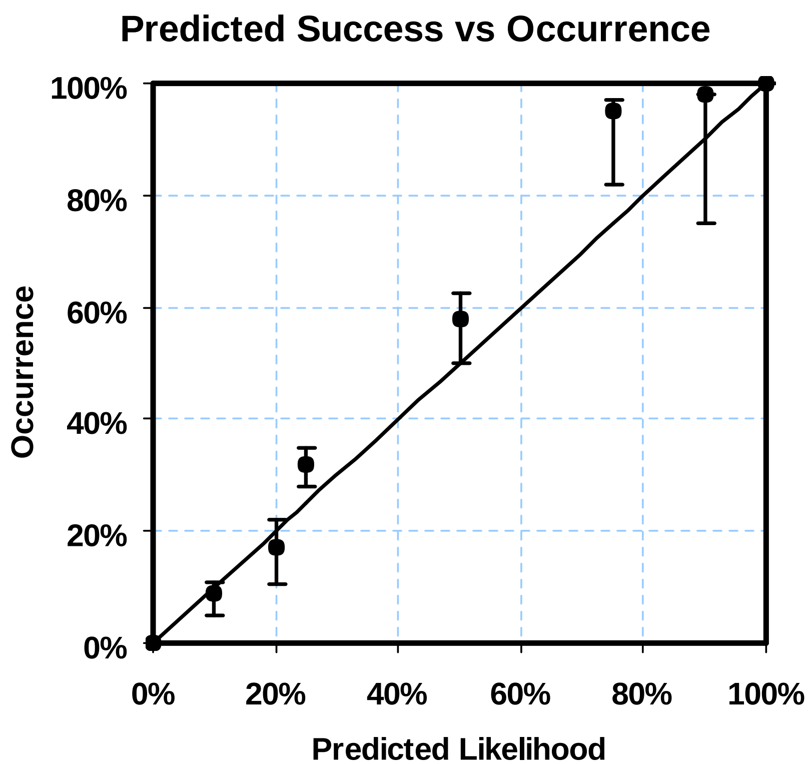
6.3.3 SUBJECTIVE FORECASTING

All forecasts are, in some respect, subjective in that the assumptions that they are based on are selected subjectively by the analyst. However, when a likelihood or confidence condition is imposed, we usually refer to the forecast as being subjective. This introduces the concept of a subjective probability. Typically, these estimates do not specifically correspond to the concept of occurrences in that there is usually only a single event that takes place. The estimate of probability here, is more of a confidence level. There are two sets of conditions where subjective probabilities are used: (1) repeated events, and (2) single occurred rare events.

[7] *Microsoft EXCEL* does not readily construct this type of analysis. However, standard statistical packages such as *SYSTAT* provides a number of tools for these procedures including confidence ellipses and kernals.

6.3.3.1 Repeated Events

Repeated events consist where estimates are made of an event for which forecasters or respondents have experience. These include sales, weather and some classes of new products (pharmaceutical testing) forecasts. Below is a comparison between subjective probability estimates and success rates for groups of ethical pharmaceuticals.



The bars around the points represent the statistical error based on the number of responds at each of the predicted likelihood levels. Notice that there is a monotonic relationship between predicted likelihood and actual occurrence. This strongly supports the value of using subjective estimates when there is a learning process due to repeated events. However, there appears to be a systematic underestimate of the likelihood at high values and a low estimate (less obvious) at low values. This corresponds to the expected bias with confidence measures. With sufficient data, subjective confidence estimates can be adjusted for use with stochastic value models.

6.3.3.2 "Rare Events"

Rare event situations are problematic since the estimators have never seen the situation actually occurring. This becomes particularly dangerous with stochastic value modeling such as with Decision and Risk Analysis (D&RA) since these probabilities are taken as the likelihood of actual occurrences. Typically, in order to make the process more reliable, the rare event situation is viewed as a series of conditional events. The conditions selected are usually similar to actual historical occurrences. The final result or joint probability of the events taking place is then compiled analytically[8]. However, in all cases this type analysis should be considered only exploratory[9].

6.3.4 ANALYTICAL EXTRAPOLATION

Extrapolating historical data is based on continuation of long term trends. There are three key elements of this type of analysis: (1) time series analysis, (2) seasonality and cyclic corrections, and (3) penetration and growth modeling. The following is merely a quick overview of the procedures.

6.3.4.1 Time Series Extrapolation

The time series analysis is designed to give the best estimate of future behavior solely based on past performance. The standard analysis is referred to as Auto-Regression-Moving-Average (ARMA) methods.[10] The auto regression part refers measuring the effect of time lags and cycles on the forecast. This is measured by some type of auto-correlation calculation[11]. The moving average computation attempts to smooth data by testing various time grouping of data.

6.3.4.2 Seasonality and Smoothing

While ARMA methods are generally used to explore periodicity and smoothing requirements, in many cases, such effects are known from the basic nature of the business. Seasonality of some businesses is notorious. Both Agricultural Products and children's toys are known to have effectively only one season a year. Petroleum products sales are also highly seasonal as would be expected. Furthermore, some sales data are inherently noisy as there is random payment of invoices and deliveries imposed upon a fixed calendar. Moving averages tends reduce these accounting artifacts.

6.3.4.3 Penetration and Growth Models

However, the most important forecasting tool is the underlying penetration and growth model used to project sales. While there are many variations on growth models there are four are generally used for long term forecasts:

---

[8] The process of merging these estimates is often referred to as Bayesian Analysis since it utilizes "Bayes' Theorem" to estimate joint probabilities.

[9] Irv Gross (formally of Dupont and Penn State) has suggested an alternative by using the game theory approaches such as "minimum regret." This has been used for pricing models and is discussed in that chapter of these notes.

[10] *Microsoft EXCEL* provides some simple ARMA methods particularly Fourier and moving average analyses. However, far more specified tools associated with time series analysis and the "Box Jenkins" methods are available on statistical packages such as *SYSTAT*.

[11] There are several type of calculations that can be performed including: (1) simple autocorrelation., (2) fourier analysis, and (3) power analysis. Results of these all show the importance of various time lags on forecasted values.

6.3.4.3.1 Linear Growth

Linear growth is typically not used unless there are small marginal changes in sales or when an extraordinary event is expected. This model is motivated from a belief that the best estimate of future sales is the present sales[12].

6.3.4.3.2 Exponential Growth

The simple exponential is the widest used trend extrapolation model. It assumes continued constant percentage growth.

6.3.4.3.3 Logistics Growth

The logistics model gives an S-Shaped growth pattern with an introduction, a growth phase, and finally a leveling to maturity. This is also referred to as a life cycle model. However, the logistics model is motivated from simple chemical processes and assumes that the mechanism for maturity is directly coupled with growth. The major problem with this model is the inability to estimate the final plateau without extensive data close to maturity. This logistic approach will be discussed in detail in the section on forecasting models.

6.3.4.3.4 The General Sales Growth Curve

The General Sales Growth Curve is a proprietary model that describes the growth phase of the product life cycle. It is discussed in great detail below in the section on forecasting models.

6.3.5 CO-PREDICTIVE FORECASTS

The most widely used forecasting procedures relate sales or earnings to a number of drivers. These are usually linear or quasi-linear regression models[13]. These procedures include general econometric models and analyzes as well as other leading indicators. The purpose of this section is to highlight the general approaches but not to review the literature on the extensive subject of econometric analysis. This is done in detail an any number of traditional references. Similar to the trend analysis used in time series analysis, co-predictor analysis uses modified multilinear regression with lead and lags in to provide the "best" forecast[14].

---

[12] This concept comes out of stock price theorem and is referred to as a "Martingale process." This results in a highly autocorrelated future price or volume with the present situation.

[13] Quasi-linear models include polynomial series and logarithmic functions. However, in this case we also include several "standard" non-linear forms such as logit and probit type regression procedures.

[14] Since several elements of a series may enter into the regression model, the number of parameters is often far larger than simple multi-linear regression and have corresponding poorer adjusted variance explained.

6.3.5.1 Early Warning Indicators

A key purpose for forecasting is to indicate demand changes as early as possible. The model used for this purpose consists of "leading" indicators of change. Unfortunately, it is usually not easy to find consistent leading indicators. Most indicators are not consistent in leading the market. Unless a factor is structurally forced to lead the market, it may vary in its effect depending on other market conditions. As such, under some conditions, a factor may be a leading, concurrent, or lagging indicator of future business activities. In cases where there are structurally forced factors such as new housing starts driving home construction supply industries, the forecast is straightforward. However, in other cases, such as forecasting movie attendance, the model is much more complex. Alternatively forecasts of concurrent factors can be used as leading indicators. However, the value of these models depends heavily on the reliability of the prior forecasts. Unfortunately, this often becomes a case of "the blind leading the blind" or worse of "wishful prophesy."

6.3.5.2 Economic Measures

The classic method of forecasting is based on general economic activities. The underlying principle is that economic and business conditions are "cyclic" in that periodically, conditions change based on underlying forces. The market responds to these forces in a consistent manner. Models are statistically constructed to relate these "macro-economic" effects to specific business activities. This approach is driven by forecasts of these general economic activities usually by a constrained historical trend extrapolation. Once again, the reliability of these forecasts rest on the prior forecasts and the identification of leading economic indicators. Unfortunately, these prior forecasts are usually unreliable[15].

6.3.6 RISK AND ERROR ANALYSIS

Understanding the underlying uncertainty in forecasts is often more valuable than the forecast itself. By their basic nature, all forecasts are uncertain. The future is not really our to see clearly! At best, we see it through the perception of history. Some effort is usually needed to estimate the degree of uncertainty. Typically there are three inherent sources of modeling forecast error beyond experimental error and bias. These are (1) uncertainty in timing, (2) errors in estimating model parameters, and (3) deviations from the fundamental model assumptions.

6.3.6.1 Time Uncertainty

Several models including the General Sales Growth Curve, which is discussed in detail later, uses the initial commercialization date as a key parameter. Experience has indicated that the timing of changes in the business situation tends to have significant uncertainty. This is particularly the situation with new businesses where the actual date

[15] It is interesting to note that often the most reliable estimate of future economic activities has been obtained by averaging the various estimates available.

of commercialization is rarely that which is either declared or even apparent. Both early introduction during development and false starts produce errors in estimating the actual initial start of new businesses.

The population balance[16]excluding dispersion is:

$$g(x,t)\text{õ}\ \text{ô}P/\text{ô}\ t - \text{ô}P/\text{ô}\ x = 0$$

where $g(x,t)$ is the forecasting growth rate model.

The boundary and initial conditions are standard for stochastic modeling:

$$\int_0^\infty P dx \;\; \boldsymbol{= 1}$$

which is simply defining the normalized probability distribution and

$$P(x,t_o) = P_o(x)$$

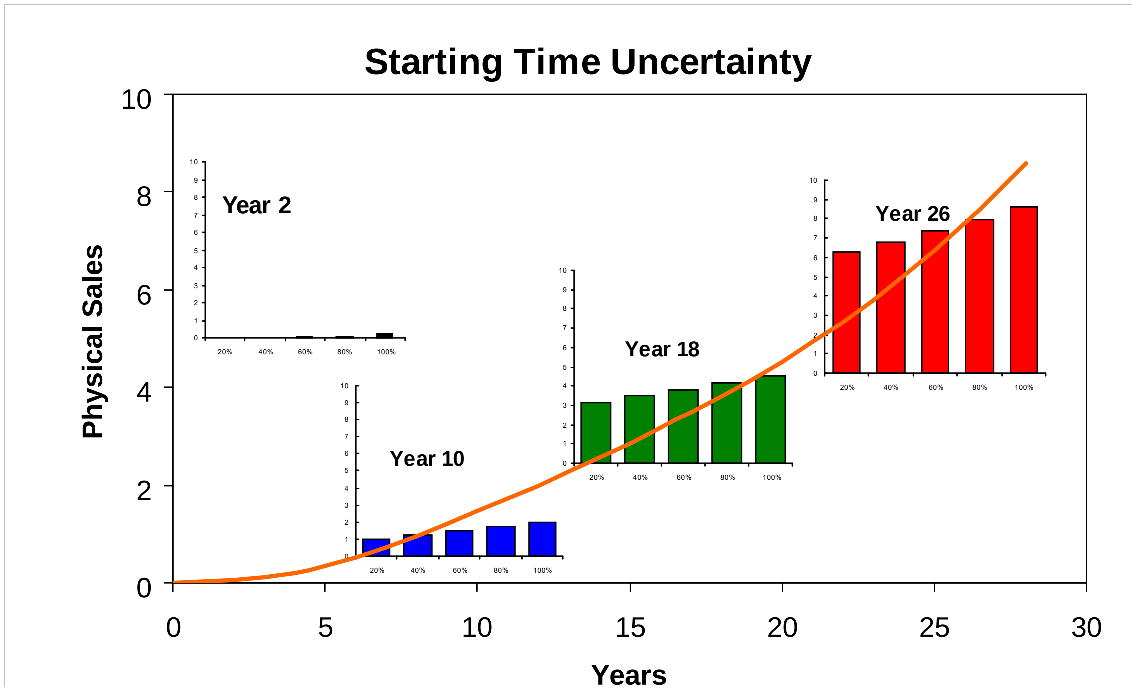
which defines the uncertainty over the initial sales volumes.

We typically use the General Sales Growth Curve which is a function only of time, $g(x,t) = g(t)$. With this assumption and under the above listed conditions, the population balance is a simple first order hyperbolic equation whose solution is a scaled time displacement. The solution is obtained by considering the initial distribution to really be over time not over sales and then recomputing new values by a time displacement.

$$P(x, t) = P_o\,(g(t))$$

This makes the estimation of the probability distribution due to uncertainty regarding timing a scaled value of the expected sales. The probability distribution of selected points in term is then associated with the corresponding sales as shown below.

[16] This relation is a simplification of the forward form of the Kolmogorov equation which describes simple continuous Markov processes [William Feller, *"An Introduction to Probablility Theory and Its Applications*", Volume 2, John Wiley (1971)].

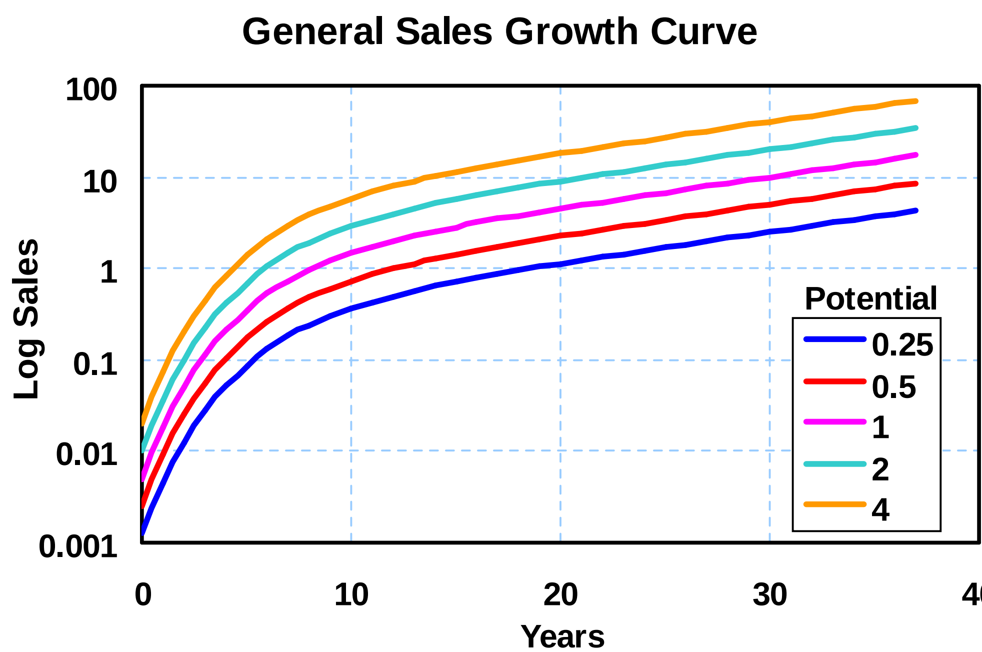


6.3.6.2 Parameter Uncertainty

The second major source of error in analytical forecasting is the uncertainty in estimating parameters. Typically variation in the parameters is handled by using various scenarios or by stochastic, probabilistic simulation[17]. The impact of that uncertainty varies widely depending on the nature of the model and the quality of fit from historical data. Experience with the General Sales Growth Curve has indicated only weak sensitivity with the only non-time adjustable parameter (market potential) compared to the variation due to timing. The impact in variation in potential for the General Sales Growth Curve is shown below[18].

[17] This is usually some form of Monte Carlo simulation.

[18] It should be noted that variation in changes of potential for this model is a vertical shift in the penetration curve on semi-logarithmic plots.



However, in some cases, such as general logistic model, forecast variations in parameter estimates have a large impact on the forecasts. This is one of the major disadvantages in the use of that model.

6.3.6.3 Aberrations from the Model

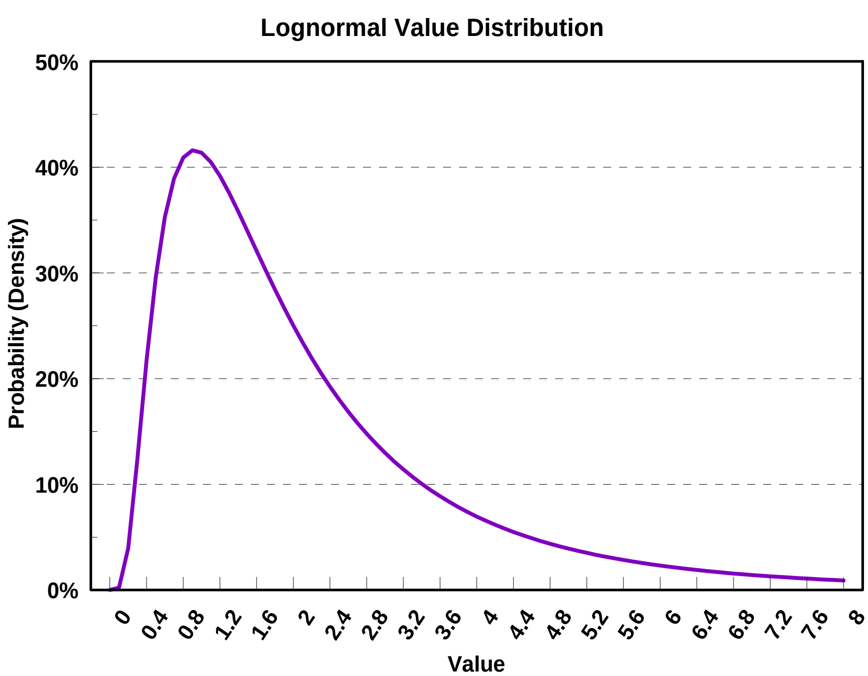
All models are based on underlying assumptions. The key assumptions are that the forces that determine growth will continue. In the case of the General Sales Growth Curve, it is assumed that the drivers of growth are maintained. However, it is recognized that all products mature and many will eventually be substituted by more cost effective technologies. When products mature, they are expected to break for the long term pattern. This is not covered, usually by growth models themselves and needs to be imposed. This is usually handled as in the case of variation in parameters by either a set of scenarios or by stochastic, probabilistic, modeling.

6.4 MARKET AND EQUILIBRIUM ANALYSIS

Market analysis is based on examining the published and channel information on competitive products to determine competitive intelligence. There is a wealth of data available by observing the market and the activities of competitor in position and promoting their products. The objectives of market analysis are to glean the effective position of the product and to identify changes in the offerings.

6.4.1 PRICE POINT DISTRIBUTIONS

The price distribution of a category of products is referred to as its price-points. This is traditionally based on merchandising of consumer package products but has been extended to consider any end-use product category. For broad product categories the distributions tend to be log-normal[19]. That is, the result curve based on the logarithm of price is described by a "normal" or Gaussian distribution. A continuous version of the curve is shown below[20].



It should be noted that is curve covers a broad category of products not the demand for a single product. It is, therefore, not the classic demand curve as has been sometimes

---

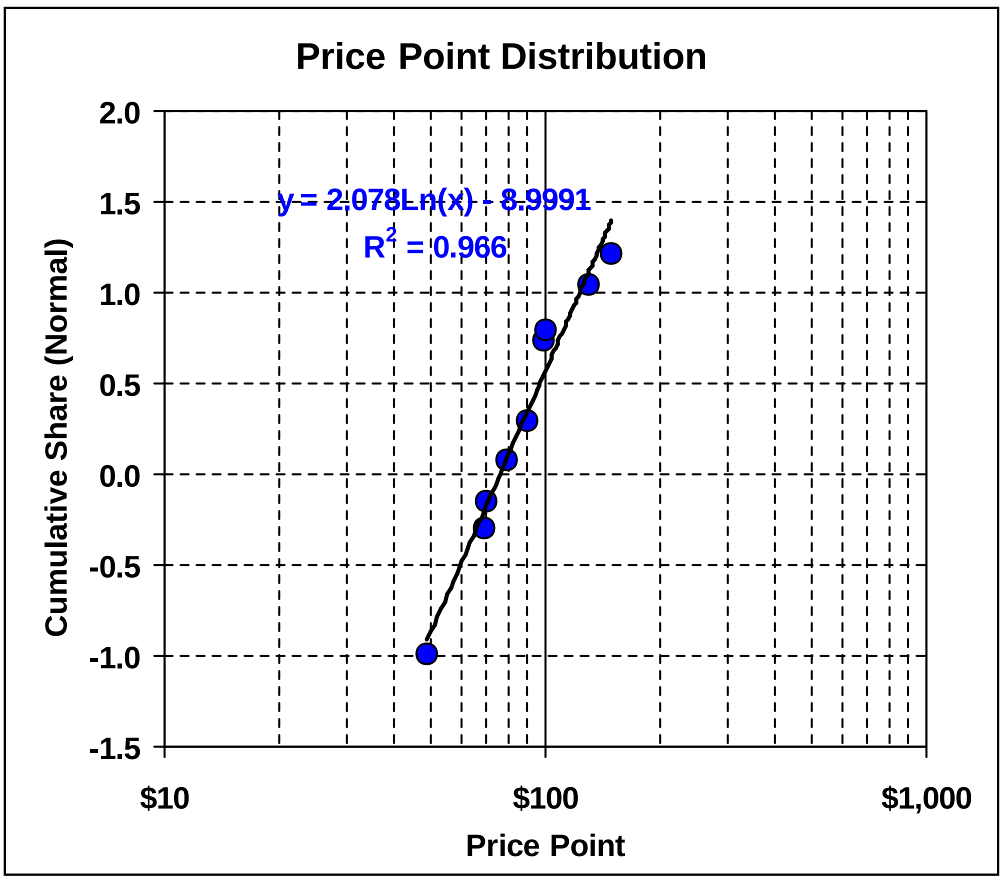
19 There are a number of similar skewed distributions that have been used to describe price-point distributions including the Weibul Distribution (Ralph Beaman formerly of Dupont). However, these usually require more parameters and do not give a significantly better fit.

20 There are several mechanisms that have been proposed that will give rise to the log-normal distribution. Most are based on assuming that customers' perceive monetary values in proportional units which corresponds to the logarithm of price. This follows Webber's law of perception which relates response to the logarithm of stimulus.

asserted.

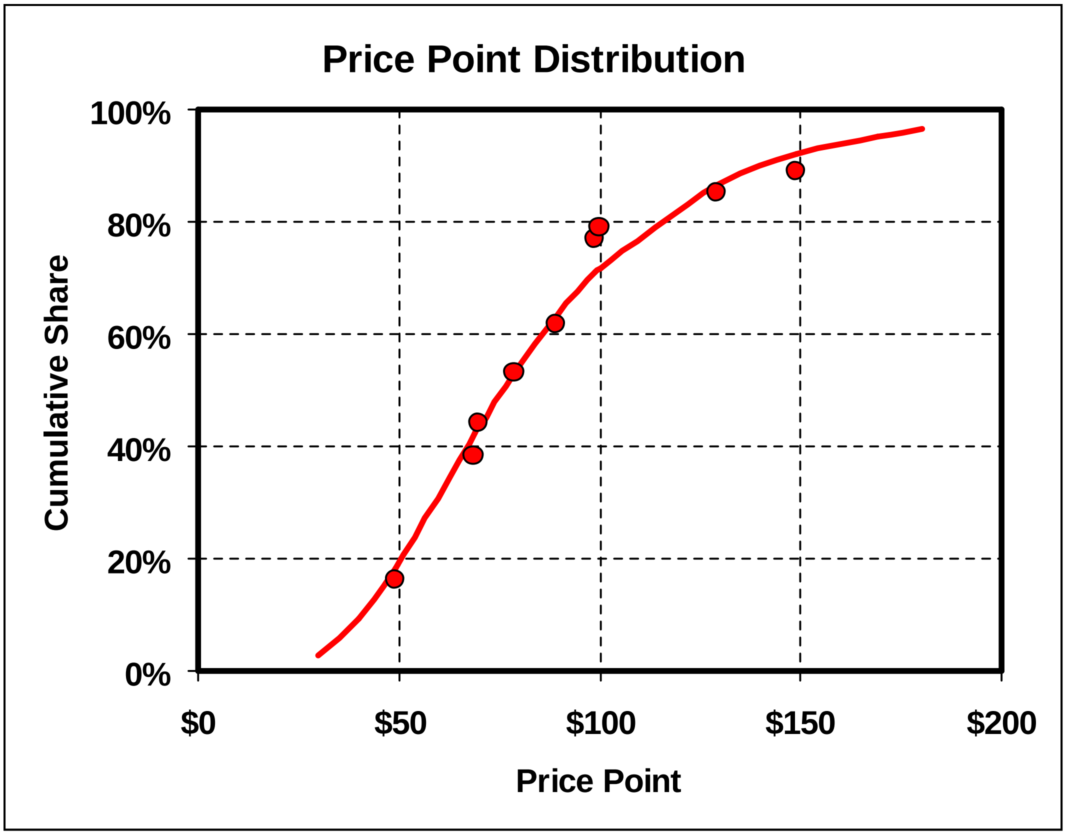
6.4.2 ADVERTISING AND SHELVE SPACE

A typical method of estimating price point distribution information is by examining advertising and store shelf data. A standard assumption, and generally confirmed experimentally, is that advertising space, such as in catalogues, or store shelf space of products is roughly proportional to the revenue derived from them. Knowing the price of the products, one can then compute the relative market shares needed to estimate the price point distribution. The data is usually analyzed graphically by plotting the price against the Gaussian of the cumulative share[21]. An example of this analysis is shown below. The curve fit is based on the appropriate regression model.



The same results are shown below in a more conventional form. Notice the errors around the curve are somewhat larger in this form compared to the graph above. This is an artifact of the conversion of the data.

[21] This can be done in *Microsoft EXCEL* which has conversion functions for the Standard Normal (Guassian) Distribution, and graphing options for logarithmic scales and appropriate regression analysis.



6.4.3 LABEL ANALYSES

Label analyses focuses on the investigation of the content and published performance of competing products. This is done based on published labels and literature. This includes published product information sheets that describe the products and the modes by which it should be used. It should be noted that this type analysis is based on data provided usually by the manufacturer. The specifications, as described, may not reflect the actual tolerances of the product. The content information is often less than totally revealing of the actual makeup of the product. Materials are often lumped together to provide manufacturing flexibility as well as to hide information. Most labels, for example, do not give quantitative measures of constituents. Usually laboratory analysis is necessary for reliable competitive product characteristication.

Use specifications, as indicated by the label and product literature, are often structured for legal considerations rather than how the product is actually used. Only where misapplication of the product is legally prohibited, such as with the commercial use of herbicides, can this information be considered reliable. In some cases, off-label use is not only expected but planned, such as with pharmaceuticals.

6.4.4 EQUILIBRIUM ANALYSIS

Equilibrium[22] analysis is based on the stability of the market. The underlying assumption is that the forces in the marketplace drive operational characteristics to stable comparative levels. The tendency of systems to move to the average, "regression to the mean," is the driving principle for this analysis. By understanding the stable levels, a forecast of future trend should be able to be discerned. There are several fundamental problems with this type of analysis including:

- The marketplace is always dynamic and "equilibrium" is almost always never achieved.
- The "Devil is in the Details" - small differences between businesses can produce huge deviations from what "should be" expected.
- Statistical analysis indicates correlation not causation. Even the most overwhelming cases may only be a result of sampling.

6.4.4.1 Herdonic Pricing

From an economic perspective the price of commodities is determined by supply and demand. The price of a product is therefore a function of scarcity. However, from a utility perspective the price that the market is willing to pay should be related to the value of its attributes. Regression analysis over a class of products should then be able to reveal the underlying values of attributes. This is referred to as "Herdonic Pricing." It is analogous to an uncontrolled experimental perceived value measurement[23].

Unfortunately, the results from this type of analysis are often disappointing. A major problem here is that the "independent" characteristics tend to be highly intercorrelated making it difficult, if not impossible, to discriminate characteristics that are not fundamentally connected. Furthermore, single highly valued products in the class can strongly affect the results in ways that would appear to be illogical. For example, Herdonic pricing of metals seems to give high value to density because gold is included. If gold is excluded, density no longer appears as a valued attribute.

6.4.4.2 Macro-Economic Factors

Major macro-economic sector spendings do not change radically over relative moderate time periods though very long term effects can be significant. For example the share of national income spent on food or education has not varied over 10% within the past decade though over a half century there has been great changes. When end-use purchases

[22] Equilibrium economic theory while inspired by the concept of thermodynamic equilibrium is really based on "steady state" situations. As such, the thermodynamic or physical equilibrium is only a metaphor for this type of analysis. This can be critical in that the metaphor has occasionally been extrapolated too far.

[23] This is structurally similar to full profile conjoint but without a statistically designed experimental basis. Full profile conjoint measurement is discussed in the chapter on perceived value measurement in these notes.

drive industry sales, the stability of these sectors spendings can be used to provide equilibrium estimates of business performance. However, these estimates should not be used for very long term forecasts.

6.4.4.3 Market Share

Volume and purchase shares in most markets tend to be consistent. Examining a broad range of markets and situations has shown that specific rank order statistical distributions can be used to estimate the eventual market shares. This approach is discussed in more detail in the section on forecasting models.

6.4.4.4 "Look-Alike" Analysis and PAR Models

When extensive business characteristic databases exist, models forecasting expected performance have been constructed. The PIMS[24] database is a collection of operational and structural characteristics of business units mainly within large corporations. There are generally two types of models that have been constructed with this data: (1) "Look-Alike" analysis and (Statistical or PAR model). The "Look-Alike" analysis consists of examining the average of a sample of businesses similar to the one under examination. A series of variations is also usually developed to explore potential strategic changes. The PAR model is a more robust analysis where a "key driver" model is constructed statistically to allow for examination to determine effective strategy.

6.4.4.5 Constrained Distribution Models

When a product group is considered truly a commodity, then the only differentiation among members is by price. Furthermore, when differences in price are linked directly to cost, then variation in manufacturing and transportation costs dominate the market. Global models based on these principles have been developed to determine likely markets and shares[25].

[24] PIMS (Profit Impact of Marketing Strategy) is a database of over 5000 businesses globally (though mainly in the US) setup during the 1960's and 70's by the Strategic Planning Institute. Unfortunately, the database has not been maintained and it is basically out of date but still useful.

[25] Several models of this type have been developed based on quadratic (mathematical) programming structures by John B. Frey (formerly of Dupont). The resulting models reproduced the historical distribution of specific commodities.

6.5 TRANSACTION AND SALES ANALYSIS, DATA MINING

Transaction data can be a wellspring of valuable information. The analysis of this type of information may include:

- Identifying key driver of purchases;
- Characterizing preferred customers;
- Identifying opportunities for product expansion; and
- Revealing underlying relationships among the variables.

Data mining refers to analyses of large databases. These databases may contain sales, operational, macro economic, market or marketing research data. These may be extremely large as in the case of transaction information, such as market scanner data or warranty cards, or in the case of employee information. They also are likely to contain a very large number of variables. For example in the case transaction data, thousands of potential products may be involved. In addition, the data is often “sparse” in that there is a significant amount of missing data. Few or any of the records contain responses to all of the variables.

These conditions limit the effective tools available for analysis. The available tools are those that require minimum memory and computation requirements. Methods that rely on computing interrelation distances (Multiple Dimensional Scaling and Hierarchical Cluster Analysis) tend to be far too memory intensive. Methods that require multiple interactions (non-linear regression, i.e. Logit and Ridge regression) are far too computationally intensive.

Because of the size of the databases and the number of variables available, it is usually necessary to limit the analysis to specific ranges of information being sought. While some work on general analyses have been done, these are usually both too memory and computationally intensive wide use, even with super-computers.

6.5.1 STATISTICAL ANALYSIS

The purpose of statistical analysis is to determine key drivers of behavior. The most effective statistical methods are based on analysis of intercorrelation. These include linear regression and factor analysis. Both sets of methods can be done based on the correlation matrix as opposed to the raw data. This involves using “pairwise” computed correlations[26]. The large number of records and the inherent problem of missing data is eliminated[27]. However, the potential impact and introduced error of the missing data is not.

[26] The general set of tools for these methods are included in the “General Lognormal Models” available with *SYSTAT*, *SPSS*, and *SAS*.

[27] This is equivalent to using the “EM Algorithm” for estimating the values of the missing data. However, that analysis does not have to be performed.

The problems using these methods include:

- There is an indeterminacy of the "goodness of fit" due to missing data. The R-squared measure (percent variance explained) requires an estimate of the equivalent number of records for the complete data which must be arbitrary when there are missing data elements.
- There are often errors in the resulting parameter values due to the use of very sparse data. Relationships may appear only because of the limited data that is available which are shown apparent strong intercorrelations.
- Potential variability in the values of parameters derived from regression analysis due to intercorrelations of the "independent" variables (also referred to as multicollinearity).

6.5.2 PATTERN RECOGNITION

The purpose of pattern recognition is to determine underlying relationships and to identify groups or segments among customers. There are generally three sets of pattern recognition methods heavily used for data mining: (1) "hill climbing" algorithms including "K-Means" clustering, (2) Bayesian pattern recognition and (3) neutral-net simulators, methods. All three sets of methods try to identify groups of respondents (or variables) that have common relationships compared to other combinations. They are based on learning mechanisms for improving predictions. They differ mainly in the formulation of the approach in identifying improvements in distinctions between items and groups.

The "hill climbing" methods are a set of clustering approaches which starts with an assumed locus of distinct concentrations and then improves the estimates based on the data. These are iterative methods which can become computationally intensive.

Bayesian methods are similar to the "hill climbing" algorithms but are used for identifying general relationships. These procedures start with a prior distribution or relationship usually based on summary data. Distinctions are then identified based on changes from the prior information. Once again, iterations of the data is often used.

Neuro-Nets are based on general layered linear models that allow interactions. The process involves progressive estimation of internal parameters that predict an outcome. Unlike the other procedures, the internal parameters of the model are not considered to be either unique nor representative of an underlying mechanism. They are only predictivers of behavior.

The problems using these methods include:

- Interpretation of the results can be problematic. In the case of neuro-nets, the underlying model is not interpretable. For the other pattern recognition procedures, while the determinate of distinction can be obtained, they are often not unique.

- Results may be dependent on the order of the data. This is a unique problem to the "learning" mechanism in that the results may depend on the order of the data analyzed. The results may not be unique or stable.

- These procedures tend to be "black-box" methods which do not lend themselves to easy interpretation or explanation.

- Some of these methods, in particular Neuronets, require a great deal of data to be functional.

6.5.3 PARALLEL PROCESSING

The limitations on memory and computation are, of course, transitory. As computers become more capable and faster, their ability to handle data analysis problems increases. However, the problem is that most statistical and data analysis computation increases quadratic ally with the number of variables and records. As such, the computational problem sky rockets. The alternative approach, however, is to use either "vectorization" or "parallelization" to allow multiple processors to work on the problems simultaneously. The pattern recognition methods allow for this type of computation and are, therefore a center piece in data mining research. However, it is uncertain at this time how this capability will translate into useable results.

6.5.4 RECOMMENDED PROCEDURES

In general, statistical methods and "hill climbing" clustering are preferred over the other pattern recognition methods, in particular over Neuronets. This is due mainly to the ability to better interpret results. However, in some cases, with sparse datasets, Neuronets have produced significantly more effective results than statistical methods. It is generally recommended, however, if Neuronets are used, statistical models also be computed and compared.

6.6 FORECASTING MODELS

Sales forecasting focuses on physical sales and revenue. Traditionally sales growth is divided between market growth and share changes. These depend on the nature of the competitive market. There are number of “standards of comparison” that can be used to help the forecasting process. In this section, we discuss the product life cycle concept and various tools for forecasting.

6.6.1 EQUILIBRIUM SHARE

6.6.1.1 Quantized Markets

Experience in industrial markets indicates that market position and share tend for individual companies and products to be stable. They do not tend to vary greatly over time unless there is a change in the market structure often associated with the entry or loss of competitors or a major change in the products offered. This stability is analogous to a "Quantum" or fixed state. Shares in this type of market are expected to follow a rank ordered distribution referred to as the "Broken Stick Rule" where share depends only on the number of competitors and the market position[28]. Because of the stability of this situation, the impact of competitive entry and exit can be estimated as well as the difficulty of changing market position. Analysis of this distribution suggests mechanisms that impose forces that inhibit additional market entries. Market leadership is shown to be particularly sensitive to the entry of competitors. The second market position, however, appears to be much more stable under these conditions.

6.6.1.2 The Broken Stick Rule

Physicists have identified that atomic particles tend to have discrete energy states. These discrete states are said to be quantized. Similarly, experience in industrial indicates that market position and share for companies and products also tend to be stable. They do not tend to vary greatly over time unless there is a change in the market structure often associated with the entry or loss of competitors. We refer to this situation as "Quantum Markets." Shares in this type of market are expected to follow a rank order statistical distribution where share depends only on the number of competitors and the market position.

Industrial markets tend to consist of few direct competitors. It is fairly well known that in the chemical and material industries most markets are limited to only a few competitors. Furthermore, market share positions tend to be stable. High levels of investment with strong economies of scale produce barriers of entry. Recognized relationships, service, reputation and established distribution channels all contribute to stabilizing share along with production capacity. However, that still does not explain the stability of share and position.

[28] The concept of the Broken Stick Rule and most of these analyses was done by John E. Reith formerly at Dupont.

6.6.1.3 Rank Order Statistics

Rank order statistics are general descriptions of expected share distributions given a ranking of participants[29]. One of the simplest of these distributions is based on random sampling of a fix linear measure. It is equivalent to taking a large number of fixed length sticks; randomly breaking them up into a fixed number of parts; ranking the parts from each stick and computing the average size of each ranked group. The average distribution from this process for four segments is 52%, 27.1%, 14.6% and 6.3%. We refer to this limiting rank order distribution as the "Broken Stick Rule." Kendall[30] has shown that this distribution can be computed in the form:

Share = 1/N [1/N + 1/(N-1) + õ õ õ + 1/J]

where **N** is the number of competitors and **J** is the ranking.

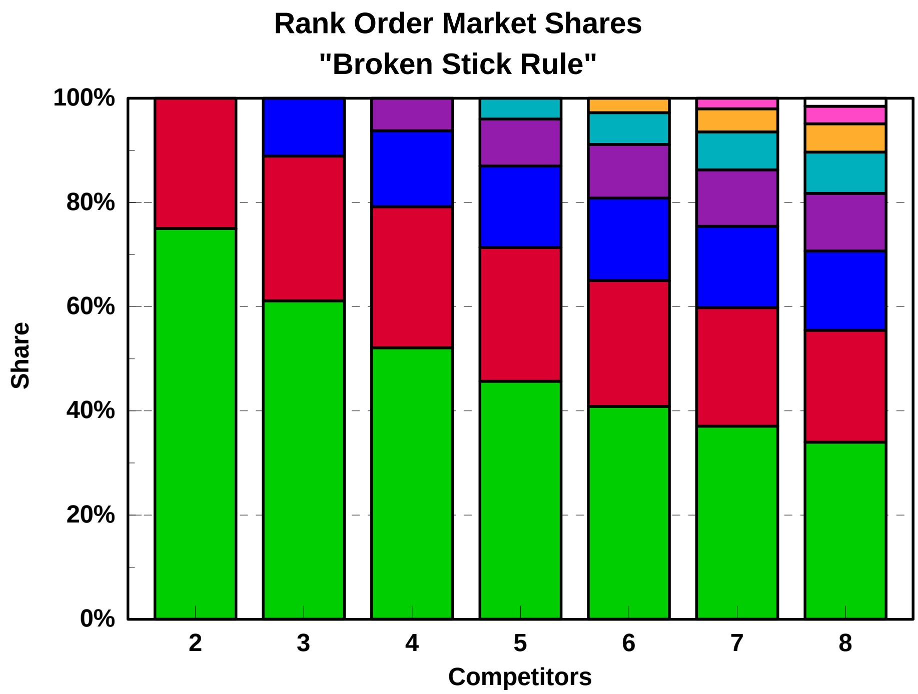
The "Broken Stick Rule" distribution appears to describe a large number of situations including value and price distributions, national expenditures by category, and industry concentration[31] as well as quantum market shares. The predicted shares are

Ranking	Number of Competitors									
	1	2	3	4	5	6	7	8	9	10
1	1.00	0.75	0.610	0.520	0.456	0.408	0.370	0.340	0.314	0.293
2		0.25	0.276	0.270	0.256	0.241	0.227	0.215	0.203	0.193
3			0.114	0.145	0.156	0.158	0.156	0.152	0.147	0.143
4				0.065	0.090	0.103	0.109	0.111	0.111	0.110
5					0.040	0.061	0.072	0.079	0.082	0.085
6						0.028	0.044	0.054	0.061	0.065
7							0.020	0.033	0.042	0.048
8								0.016	0.026	0.034
9									0.012	0.021
10										0.01

[29]Gumbel, E. J., **Statistics of Extremes**, Columbia University Press, NY pg. 42 (1958)

[30]Kendall, M. G., "Ranks and Measures", *Biometrica* (1962), 49, 1 and 2, pg. 133

[31] Kendall paper focused on psychological testing.



Based on this model, for example, with two competitors, the leader would have 75% of the market while the second would have the remaining 25%. This is intuitively reasonable. If we break the stick into two parts, the larger part will be uniformly distributed between 50% and slightly less than 100% with an average of 75%. The other cases are less obvious.

6.6.1.4 A Market Description

One of the most difficult aspects of building market share models is testing them against data. The very nature of market share makes this problematic. Share data are proprietary and are exposed to varying definitions for strategic planning. Even when the product is well structured, the geographic area of interest tends to be ill-defined.. Because of these limitations any test of market share models must be viewed as anecdotal rather than statistical[32].

Some supporting evidence for the "Broken Stick Rule" can be seen in the US automotive market in 1967, before the entry of strong foreign competition. The shares of General Motors, Ford, Chrysler and American Motors were 52%, 27%, 14.5% and 6.5%. This is in very close agreement with those predicted by the "Broken Stick Rule." An industrial example for products can be seen in the dynamic behavior of the carpet fiber market.

[32] John B. Frey showed that the market share from the SRI's PIMS database agreed in general with the Broken Stick Rule (Private Communications 1982).

The comparisons for market shares in 1963 were:

Fiber	Actual Share	Predicted
Wool	47%	46%
Nylon (Continuous)	26%	26%
Nylon (Staple)	13%	15%
Acrylic	9%	9%
Rayon	5%	4%

This is in good agreement with the predicted values. By 1971, the market had radically changed in terms of both participants and shares. New fibers were introduced along with changes in pricing. However, the "Broken Stick Rule" was still a good descriptor of the market structure.

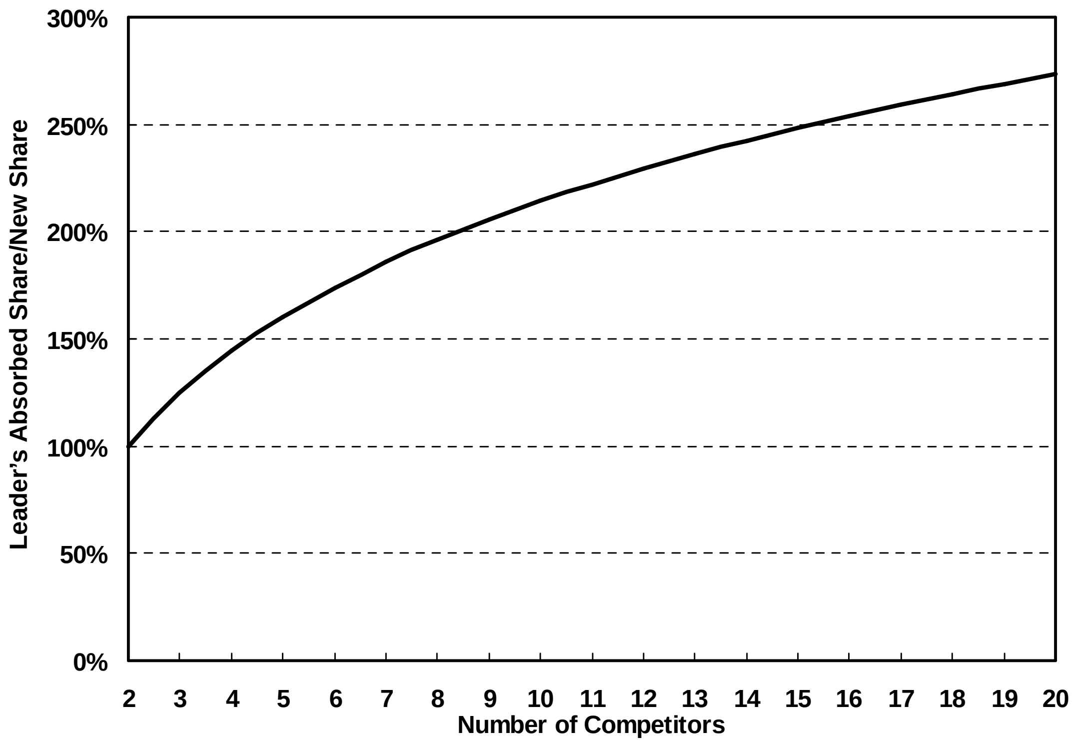
Fiber	Actual Share	Predicted
Nylon (Continuous)	34%	34%
Acrylic	24%	22%
Polyester	14%	15%
Nylon (Staple)	12%	11%
Wool	6%	8%
Cotton	5%	5%
Polypropylene	3%	3%
Rayon	2%	2%

6.6.1.5 Implications

A market is "quantized" if the market positions are stable once formed and that shares follow the "Broken Stick Rule." If the market is quantized there are important implications based on the behavior of predicted share distribution.

6.6.1.5.1 The Price of Leadership

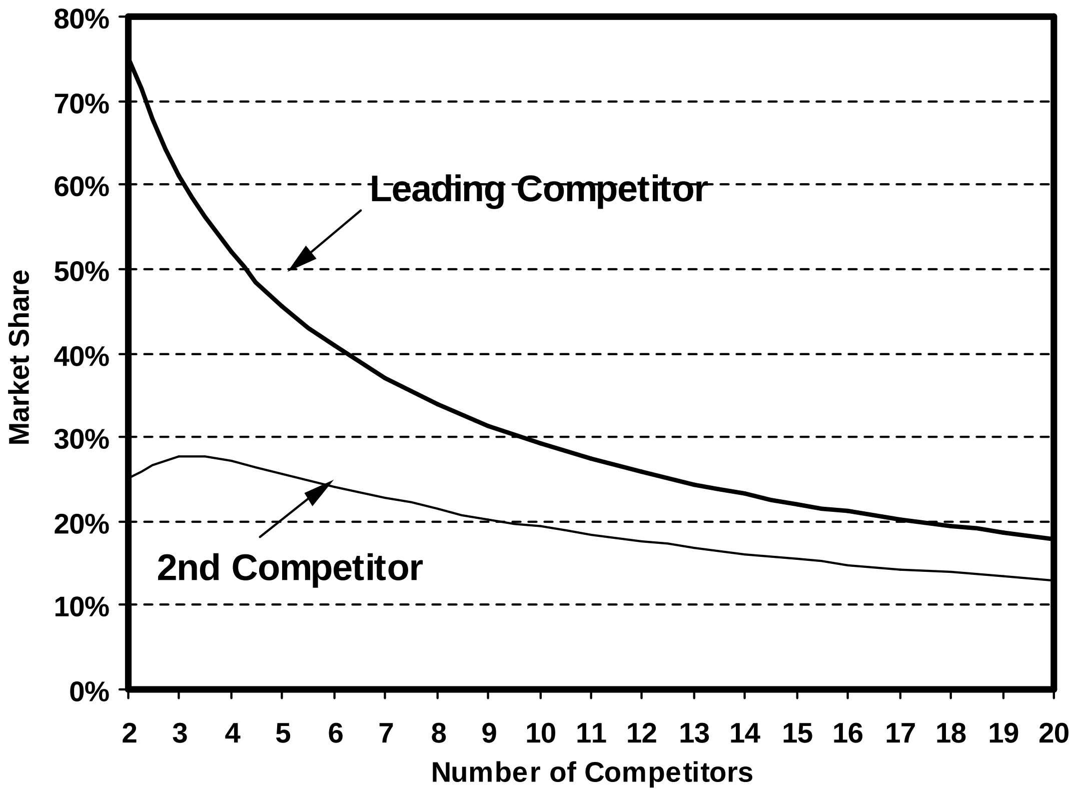
The entry or exit of a competitor results in the redistribution of sales. The new entry, for example, gains share primarily from the market. However, it appears that other minor participants also gain share. The result is that the market leader takes a major "hit" with the entry of any new competitor or gains when a competitor leaves the market. Not only does the leader lose share, but that loss is greater than the new entry's gain. The following figure shows the ratios of the loss of the leader's share divided by the gain of the new entry. Obviously, the ratio is equal to one when the first new competitor enters the market, since any gain in the new entry comes at the expense of the leader in this case. However, as new entries come on to the market the loss of the leader greatly exceeded the marginal gain by the new entry. The rest of the gain is to other intermediate competitors. It is natural, therefore, to expect the share leader in a market to resist the entry of any new competitor.



Leader's Absorbed Market Loss with New Entries

6.6.1.5.2 Stability of the Second Position

The following figure shows the market shares for the leader and the second largest competitor. Notice that while the leaders share drops-off exponentially with the number of competitors, the second largest does not. In limited competitive markets, with less than eight competitors, the second largest competitor's share varies only between 23% and 28%. The market leader, however, shows a vast change in share, ranging between 75% for two competitors and 37% with 7 competitors. The stability of the second position offers a strategic advantage with potential pricing implications in relationship with the market leader.

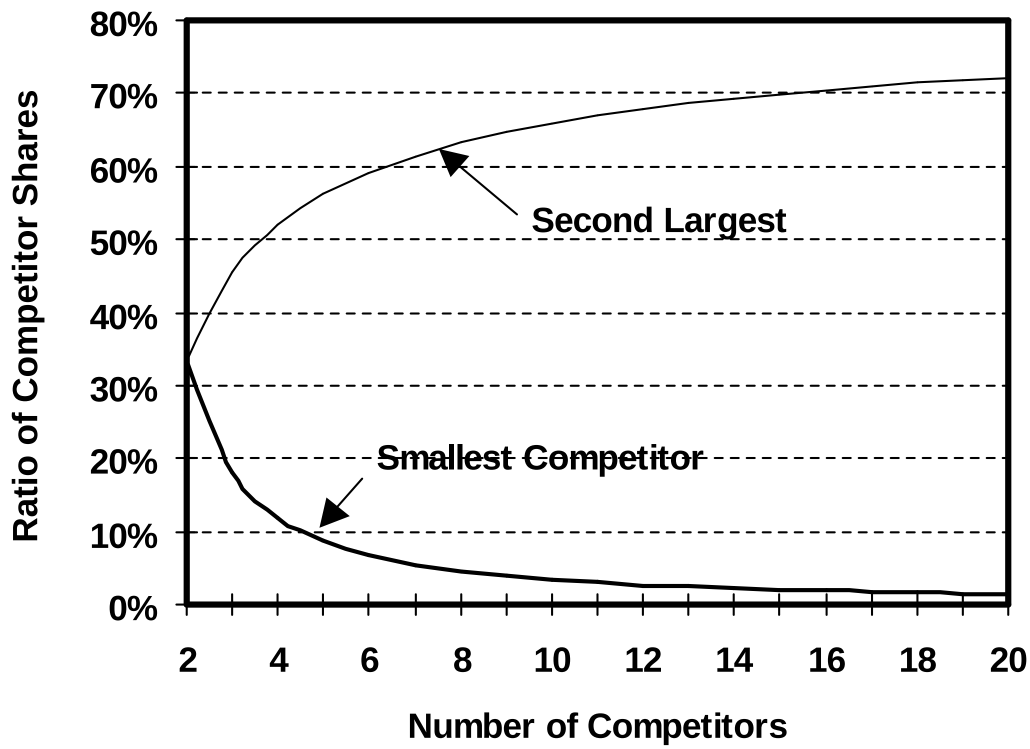


Market Share for Leading and 2nd Competitors

6.6.1.5.3 Advantage of Size

Due to any economy of scale, the market leader would enjoy some a competitive advantage over its competitors. With strong economies of scale, which are typical in many industrial businesses, this translates into a major cost advantage. The next figure shows the ratio of sizes between the second largest and the smallest competitor compared to the leader. The relative size of the smallest competitor rapidly declines making them increasingly disadvantaged. Any significant economy of scale naturally limits competitive entry. With large initial capital investment requirements, it is, therefore, not surprising that the competition in the material and chemical industries is limited to only a few competitors.

However, the situation for the second largest competitor is entirely different. The ratio of his sales to that of the leader increases with the entry of new competitors. The second largest competitor may logically encourage new market entries from a relative cost perspective. The new entries should have only a minor impact on his sales and yet improve his relative position compared to the leader.



Ratio of Market Shares of Smallest and Second Largest to the Leader

6.6.1.5.4 Stability of Shares

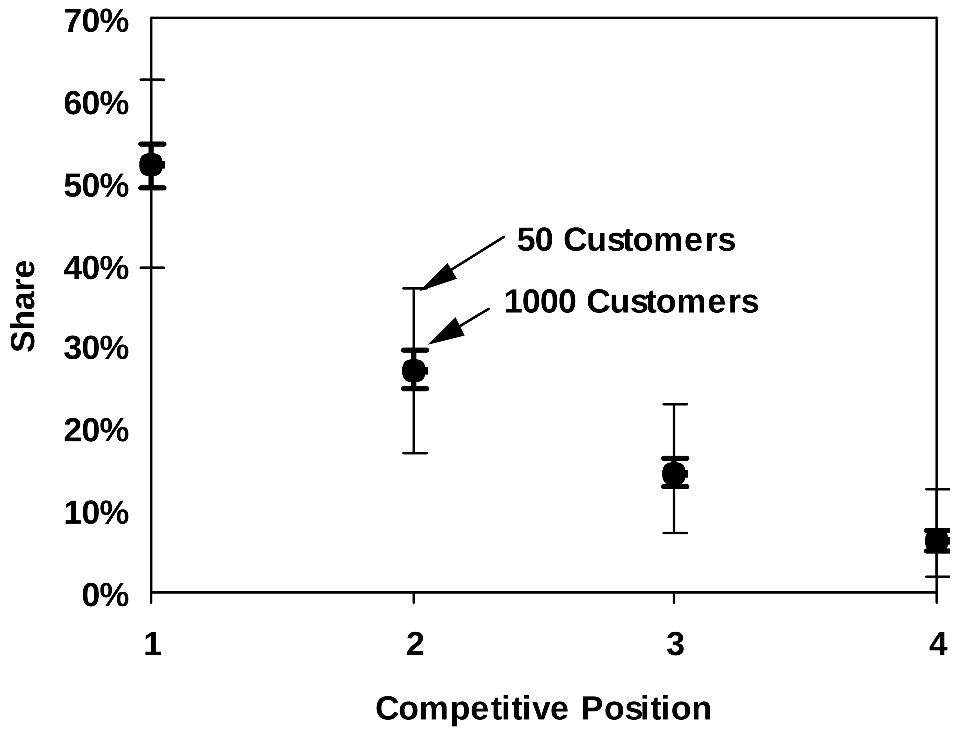
The difficulty of changing market position depends on the number of customers involved. With many customers, it is expected to be a difficult task to challenge the existing structure. The larger the number of customers, the more stable will be the shares. However, in some industrial markets, the number of customers is often quite small. In these cases, we would expect significant variation with corresponding changes in market position.

In order to estimate this variation we assume that the variation around the "Broken Stick Rule" will follow standard estimates of share[33]. The figure below shows confidence intervals[34] around the share estimates for 50 and 1000 customer markets with 4 competitors. The larger number of customers, the tighter the interval. The darker indicated interval is for 1000 customers. The broader interval is for 50 customers. Notice that with 1000 customers each of competitive share position is distinct. However, with the smaller market the shares begin to overlap..

---

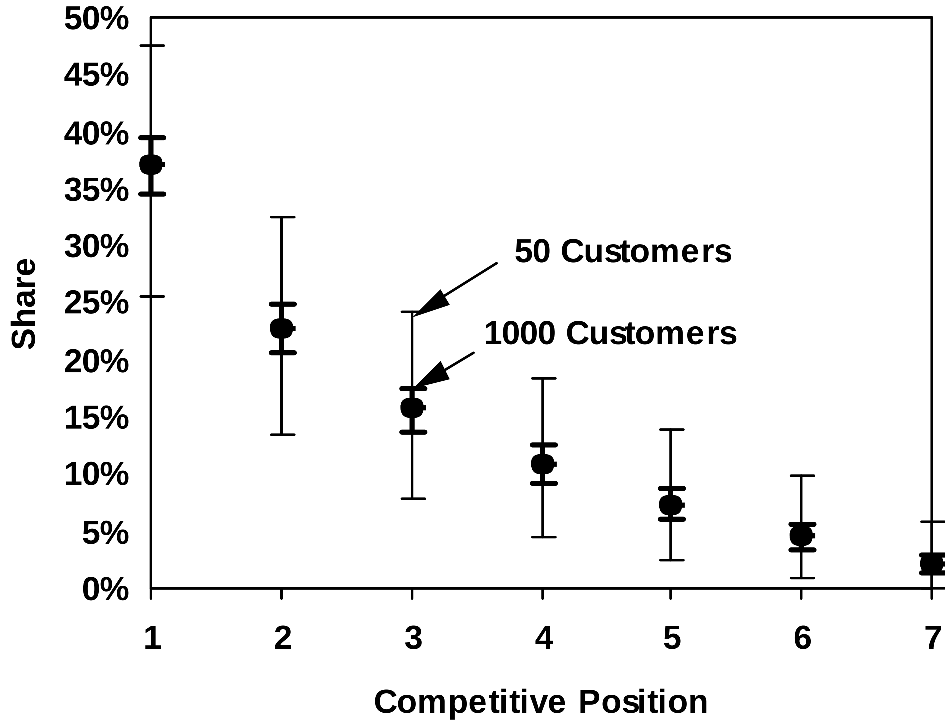
[33] This is a binomial sampling distribution with the sample size equal to the market. Feller, W. **An Introduction to Probability Theory and Its Applications**, Volume 1, John Wiley & Son, NY, pg. 146 (1968)

[34] These are 95% confidence intervals.



Confidence bound around 4 Competitors with 50 and 1000 Customers

The situation is even more marked when 7 competitors are considered as shown. Here even for the large markets of 1000 customers, the share levels overlap. Under these conditions, market positions are uncertain particularly with small markets. Only the market leader, and may be the second competitor, is market position clear in markets with few customers markets. Changes in market position should be fairly easily to achieve for lower positioned competitors in these cases. In large industrial markets, with large numbers of customers, the market shares should be stable, particularly, if there are few competitors. This is normally the case in major industrial markets.



Confidence bound around 7 Competitors with 50 and 1000 Customers

6.6.1.6 Efficient and Orderly Markets

Markets can become overly competitive. With very strong economies of scale, situations do occasionally exist where competitive pressure will periodically drive prices below average costs. Under these conditions, the industry becomes unprofitable for all participants. This is not in the interest of either the suppliers or the customers since it usually results in price instability as well as reduced services, innovation and reinvestment.

If markets are Quantized, that is that market share is inherently stable unbridled price competition will not change share significantly. These markets are orderly in that competitors' positions do not readily change and may signify an "efficient" market from an economic perspective. Since shares are more predictable, expected sales volumes should also be more predictable, improving the investment efficiency and producing lower long term costs. As such these markets could capture relatively higher margins. It is easy, therefore, to suggest that "Quantum Markets" are a preferred natural competitive structure. However, this assertion is only speculation. Future research will be needed to confirm this speculation.

6.6.2 SALES GROWTH CURVES

6.6.2.1 Potential Applications

If the physical sales of growth of products is defined by a common shaped curve then we may have a means to: (1) bound long range sales forecasts, (2) determine when a product line is likely to have matured, and (3) explore the consequence of business development policies. As part of this project, we are testing the ability of the sales growth relationship to forecast sales as well as to provide confidence bounds around the forecasts.

6.6.2.2 Background and Motivation

Research on technology substitution has indicated that sales growth follows orderly dynamics. Models based on this tradition have suggested that the dynamics of new product growth depend on the nature of the product and its application value. However, we wish to investigate if the dynamics of growth of industrial products may be independent of the type of product. The shape of the physical sales curve over time during growth might be the same for many product categories.

We suggest that this common sales behavior may extend only over the growth phase of the product life cycle. All products eventually mature, so sales will deviate from this common curve at some point. Murray [35] observed abrupt deviation from the diffusion curve when fitting product penetration data. The problem of modeling product maturation is, therefore, handled separately from that of product growth.

This preliminary report covers: (1) development of a data base for testing product sales dynamics, (2) the testing of aspects of the common sales growth curve, and (3) a discussion of some issues relating traditional substitution analysis with this research.

6.6.2.3 Models of Sales Growth

Our hypothesis is that the growth of physical sales of products follows the same dynamic function. To test this hypothesis it is useful to adopt simple analytic functions to describe sales growth. The form of this function must capture three key characteristics: (1) a potential that is unique to the product, which grows at a constant rate, (2) a non-symmetric penetration curve, and (3) a unique commercialization date.

A useful functional form for the growth curve consists of two parts: (1) potential function and (2) penetration relationship.

$$\textbf{Sales} = [\textbf{Potential}][\textbf{Penetration Fraction}]$$

$$\textbf{Sales} = P_o(1+i)^{T-T_o}\,[\textbf{Penetration Fraction}]$$

[35] Murray, S. O. & J. H. Rankin, Use Diffusion: An Extension & Critique, Tech. Fore. & Soc. Chan., 16, 331-341 (1980)

where **i** is the long term growth rate of the potential; $(T\text{-}T_o)$ is the time since commercialization and P_o is the initial market potential (at T_o).

6.6.2.3.1 The Gompertz Model[36]

A main issue in this research is the form of the penetration function. Several functions are capable of describing this behavior, including non-symmetric penetration, such as the log-normal [37]and the Weibull curves[38] . However, a Gompertz curve with a fixed penetration rate is a candidate for the simplest form. It is interesting to note that Lakhani [37] also found that the Gompertz curve gave a better fit for market penetration of new products than either the log-normal or the logistic models. The Gompertz curve with a fixed penetration rate is a two parameter penetration function or a three parameter growth model with the third parameter being the long term growth rate.

We will show that the best fit to our data gives a long term growth rate of 9.15%[*] . A model that meets this initial conditions is:

$$U = P_o(1.0915)^{T-T_o}\,(U_o/P_o)^{R^{(T-T_o)}}$$

were **U** is the physical sales over time, P_o is the first year potential which grows at 9% annually, U_o is the first year sales, and **R** is a constant (0.77). When this relationship is transformed into a linear function the three parameters combine to form two independent parameters for curve fitting.

6.6.2.3.2 The Logistic Model

Most traditional models for both sales penetration and technological substitution are based on modifications of the logistics curve or other "S-shaped" functions[39]. This function has two key characteristics: (1) an asymptotic limit, and a mirror symmetry, which gives the inflection point mid-way through the history of the penetration. Marchetti[40] has shown that this type a model gives fairly accurate forecasts for the substitution of technologies as well as other social dynamic phenomena. As a first order

---

[36] The use of Gompertz Curve for sales growth curves was suggested by Carl Jepson formerly of Dupont as was the initial motivation of this type of analysis.

[37] Lakhani, H., Empirical Implications of Mathematical Functions Used to Analyze Market Penetration of New Products, Tech. Fore. & Soc. Chan., 15 (1979)

[38] Sharif, M. N., & M. N. Islam, The Weibull Distribution as a General Model for Forecasting Technological Change, Tech. Fore. & Soc. Chan., 18, 247-256 (1980)

[*] In most of our work we still use the 8% growth which is characteristics of chemical products (1996 revision)

[39] Jantsch, E., TECHNOLOGICAL FORECASTING IN PERSPECTIVE, Organization for Economic Co-operation and Development, Paris, France (1967)

[40] Marchetti, C., Society as a Learning System: Discovery, Invention, and Innovation Cycles Revisited, Tech. Fore. & Soc. Chan., 18, 267-282 (1980)

estimate for such forecasts, this model does well.

6.6.2.3.3 Comparison of the Logistics and Gompertz Curves

The logistics model has been used to describe the substitution of one technology by another and constitutes the bulk of the literature on product penetration [41, 42]. However, investigators have noted the need to modify the logistics model to handle deviations for product penetration. These include varying long term "asymptotic limits" [43, 44, 45] similar to the type that is used in our model, and non-symmetry for the penetration of curves[46]

The Gompertz curve is another S-shaped curve, but differs from the logistics curve in that it is not symmetric. The inflection point of the "S" for the logistics model is at the 50% penetration point, while for the Gompertz curve it is much higher (two-thirds). Our research consists of testing these two and the simple constant proportional growth (exponential) models using a specially collected database..

6.6.2.4 Procedures

We have followed a research procedure with the goal of identifying new simplifying principles. The scientific paradigm [47]endeavors to develop models based on their: (1) generality in a limited environment, (2) validity compared with available data, and (3) simplicity.

Previous business, social, and economic research has tended to rely on structuralist and case study approaches. These research approaches have somewhat different objectives than the classic scientific form. The objectives of the structuralist paradigm are to construct model which: (1) describes the total history of the substitution process, (2) are robust to handle potential variations, and (3) has a clear underlying causal mechanism [48]. This approach lead to the development of complex models able to handle broad families

[41]*1* Chow, G. C., Technological Change and the Demand for Computers, Amer. Eco. Rev, 57 (5), 1117-1136 (19)

[42] Sanford, T. W. L., Trends in Experimental High-Energy Physicals, Tech. Fore. & Soc. Chan., 23, 25-40 (1983)

[43]Giese, R., P. C. Jones, & B. G. Kioetch, Electric Vehicles, Market Penetration and Positive Externalities, Tech. Fore. & Soc. Chan., 24, 137-152 (1983)

[44] Mahajan, V. & R. A. Peterson, First Purchase Diffusion Model of New Product Acceptance, Tech. Fore. & Soc. Chan. 15: 127-146 (1979)

[45] Stier, J. C., Technological Substitution in the US Pulp and Paper Industry, Tech. Fore. & Soc. Chan., 23, 237-245 (1983)

[46] Easingwood, C., V. Mahajan, V. & E. Muller, A Nonsymmetric Responding Logistic Model for Forecasting Technological Substitution, Tech. Fore. & Soc. Chan., 20, 199-213 (1981)

[47] Kuhn, T. S., THE STRUCTURE OF SCIENTIFIC REVOLUTIONS, University of Chicago, Chicago, Ill. (1962)

[48] Mitroff, I. I., The Philosophy of Modeling & Futures Research, A Guide to Different Models, Tech. Fore. & Soc. Chan., 21, 267-280 (1982)

of deviations[49, 50, 51] and to develop means of estimating multiple parameters[52, 53, 54].

Our objectives are somewhat different; we wish to verify only first order effects for limited, well specified conditions. We do not intend to propose any specific mechanism that gives rise to the observed common behavior. Rather, we seek only to describe what appears to be a rather pervasive phenomena leaving explanation of that phenomena to later work.

6.6.2.4.1 The Data Base

Our data base consists of 302 sets of time series of material category sales, each including the initial high growth period. The series include annual physical production or consumption data varying from as short as six years up to 175 years. The data are mainly of U. S. national production data for generic classes of products. There is also some foreign national data series included. Most of the data was obtained, either directly or indirectly, from government sources. We obtained the data from:

- The Chemical Economic Handbook (SRI);
- The Statistical Abstracts of the United States (From Colonial Times to 1970);
- Kline's Industry Reviews; and
- SEMI Annual Electronics Industry Report.

Missing from the database are brand sales data and detail application information. The files for the SRI data were modified by using a random scaling factor, a condition accepted by SRI in freely using and publishing their data series.

6.6.2.4.2 Testing Procedures

There are three key issues within our overall hypothesis:

H_1: Does a general (2 parameters) curve capture the majority of the variation in

---

[49] *8* Mahajan, V. & M. E. F. Schoeman, Generalized Model for the Time Pattern of the Diffusion Process, IEEE Tranaction on Engineering Management, EM-24, 12-18 (1977)

[50] Sanatani, S., Market Penetration of New Products in Segmented Populations: A System Dynamics Simulation with Fuzzy Sets, Tech. Fore. & Soc. Chan., 19, 313-329 (1981)

[51] Silverman, B. G., Market Penetration Model: Multimarket, Multitechnology, Multiattribute Technological Forecasting, Tech. Fore. & Soc. Chan., 20, 215-233 (1981)

[52] Olson, J. A., Generalized Least Squares & Maximum Likelihood Estimation of the Logistic Function for Technology Diffusion, Tech. Fore. & Soc. Chan., 21, 241 - 249 (1982)

[53] Martino, J. P., The Effect of Error in Estimating the Upper Limit of a Growth Curve, Tech. Fore. & Soc. Chan. 4: 77-84 (1972)

[54] Willis, R. E., Statistical Considerations in the Fitting of Growth Curves, Tech. Fore. & Soc. Chan., 15, 107-125 (1979)

sales during the growth phase of the product life?

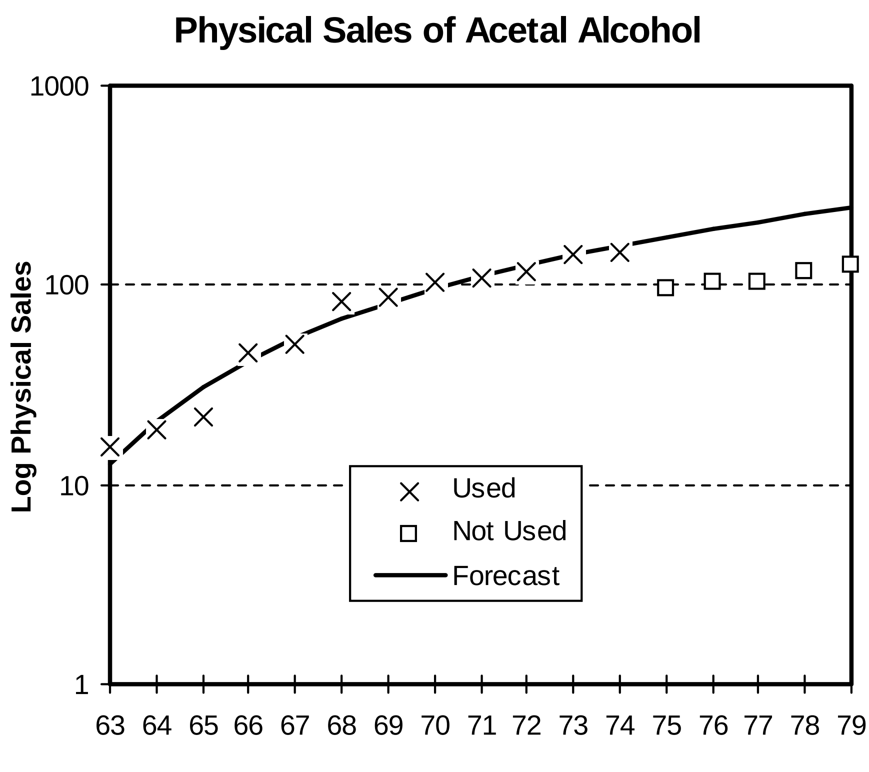
H_2: Do long term growth approaches a limiting rate?

H_3: And the deviation (maturation) from the growth curve abrupt?

These hypotheses focus exclusively on the stable growth phase of the product life cycle. In order to test them, we need to exclude sales data from both the initial startup period and maturity. Very early data (within the first two years) may have inaccuracies in reporting and extraordinary events during initial commercialization. The following set of rules were used to reject data:

- Delete early data (within the first two years) if their inclusion degraded the fit of the first six points by 30%; and
- Delete later points if a series of three subsequent points deviated from the curve by greater than 20% in the same positive or negative direction.

Models (and data rejection criteria) were tested using linear regression of appropriately transformed variables. The comparative percentage of variance explained (R-Squared) was used as the primary measure of fit. The following figures illustrate typical data series.



Typical fit of data to the 2 -Parameter Gompertz (General Sales Growth) Curve

In order to analyze the population of curve fits, we found it useful to examine the "logit" transformed R-Square measure, **-log[1 - 1/{R-Squared}]**. The set of transformed R-

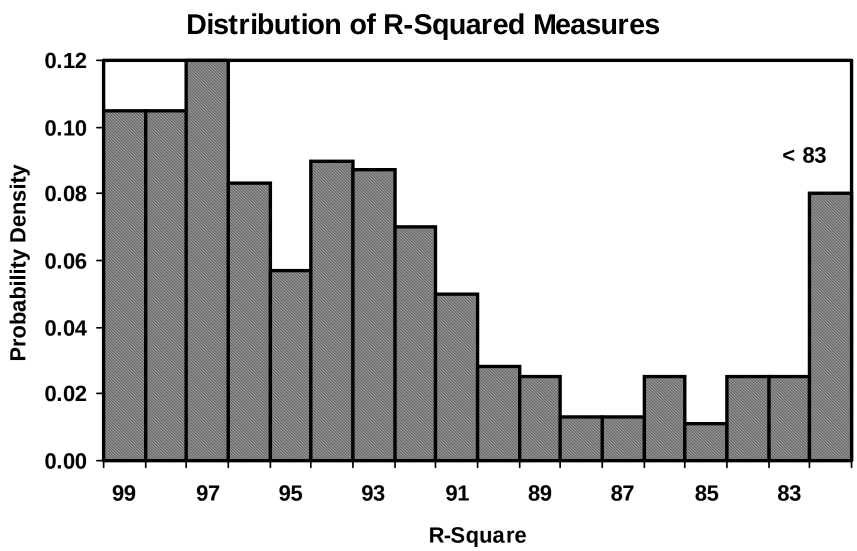
Square tended to be normally distributed (Gaussian).

6.6.2.5 Results

6.6.2.5.1 Agreement with a Common Curve

The two parameter Gompertz growth curve (with an 9.15% growth of potential and a fixed penetration rate) explained 91.4% (standard error = .1%) of the variance compared with a 74.1% (standard error = 2.3%) for the constant exponential growth model for the 302 series. . Figure 3 shows the distributions of R-Squares for the two parameter Gompertz curve.

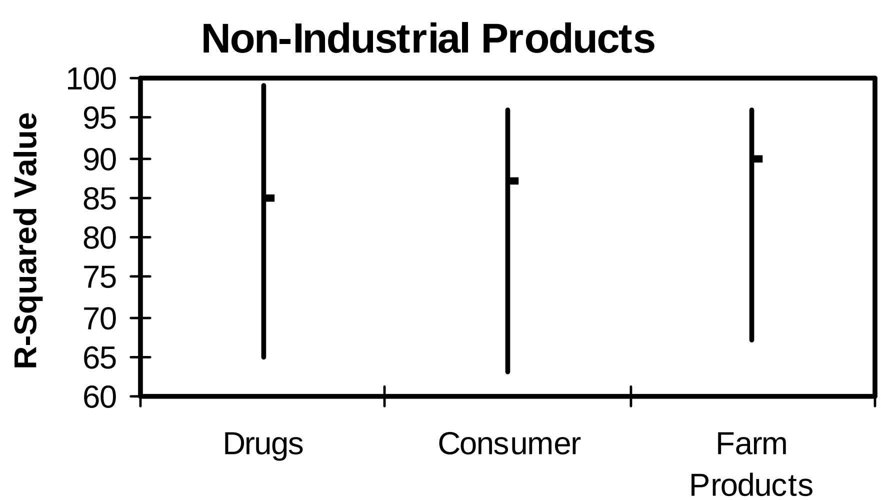
It is clear that the growth curve is a far better fit than the exponential model. This implies that that the two-parameter model can capture the vast majority of the variation of physical growth data. However, there is variation in the fit with differing product categories. Thirteen product and four geographic categories were separated from the database. The following figures show the resulting R-Square results. On those figures the ends of the bars represents the extreme range of the results. The tic marks on the bars are the mean values.



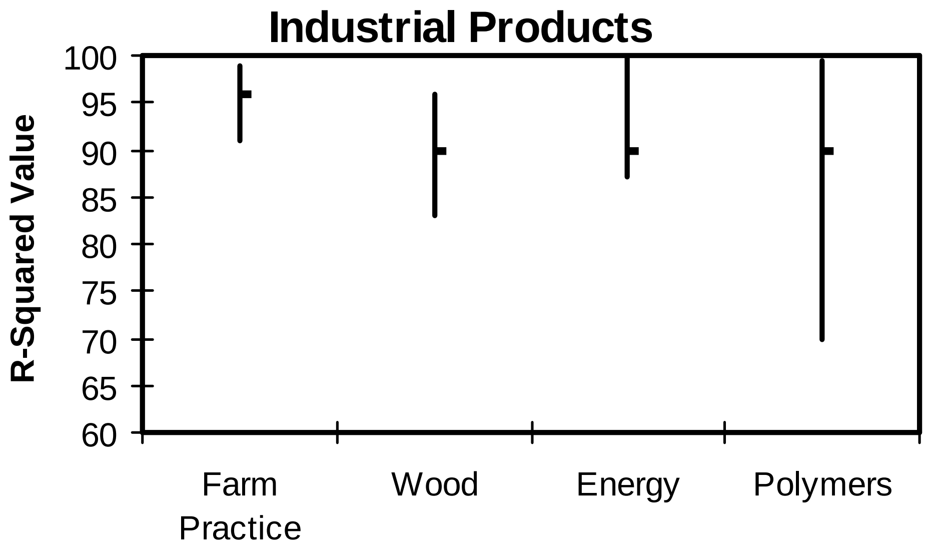
The Distribution of Percent Variance Explained of General Sales Growth Curve

	Gompertz (General Sales Growth)				Exponential	
	#	Mean R^2	Max R^2	Min R^2	#	Mean R^2
Pharmaceuticals	19	85	99.0	65	19	80
Farm Products	8	90	96.0	67	8	71
Consumer Productgs	13	87	96.0	63	11	65
Farm Practices	9	96	99.0	91	5	64
Wood Products	9	90	96.0	83	9	80
Petroleum/Energy	15	90	99.6	87	12	76
Polymers	30	90	99.5	70	30	76
Technologies	23	92	99.0	72	22	72
Chemicals	30	92	98.0	71	29	68
Explosives	4	73	98.0	56	4	80
Tire Cord	3	91	98.0	81	2	68
Amino-salts	5	83	99.0	63	5	87
Applications	26	90	99.0	68	26	82
Geographic Areas						
Western Europe	13	93	99.6	81	13	85
Japan	119	81	97.0	61	9	82
Eastern Europe	8	99	99.7	82	8	92
Developing Countries	12	93	99.7	84	11	86

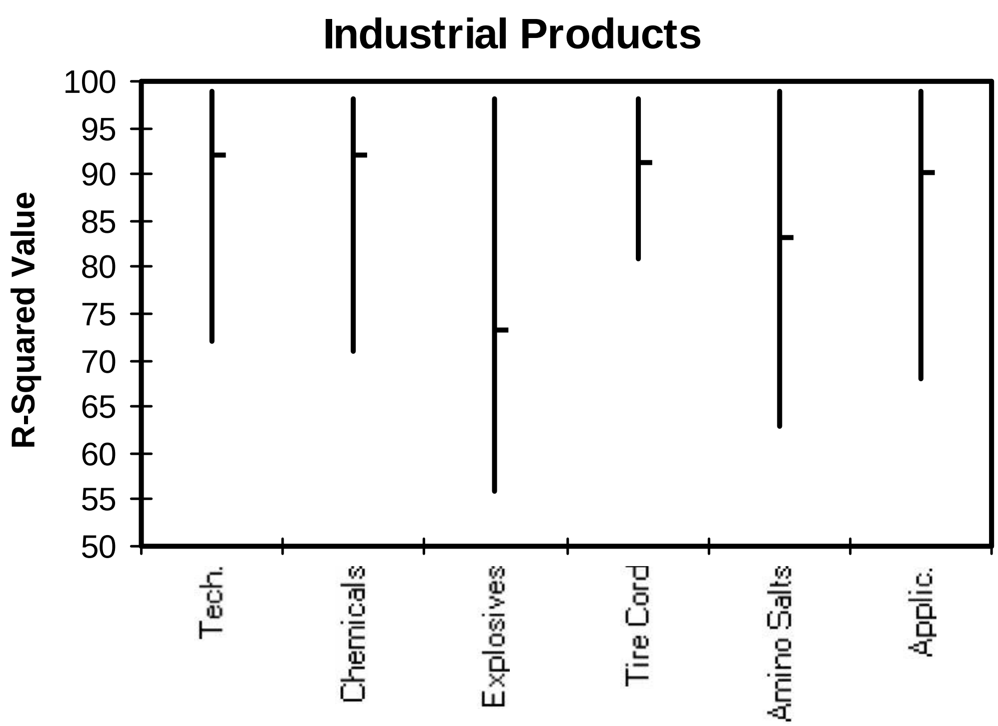
Comparison of the Fit to the Modified Gompertz vs the Exponential Curve



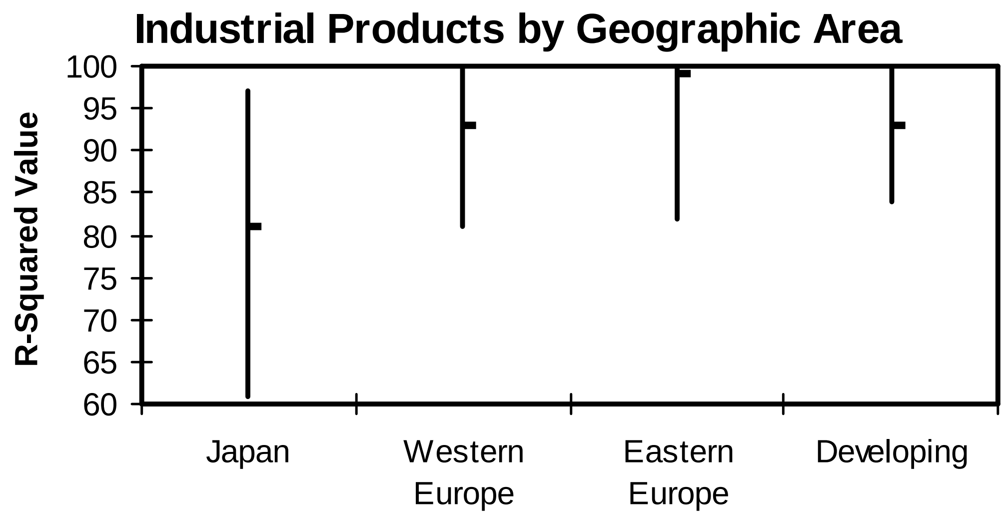
Fit of Non-Industrial Product Data Against the Modified Gompertz (General Sales Growth) Curve



Fit of Non-Industrial Product Data Against the Gompertz (General Sales Growth) Curve



Fit of Industrial Products/Technologies Against the Gompertz (General Sales Growth) Curve

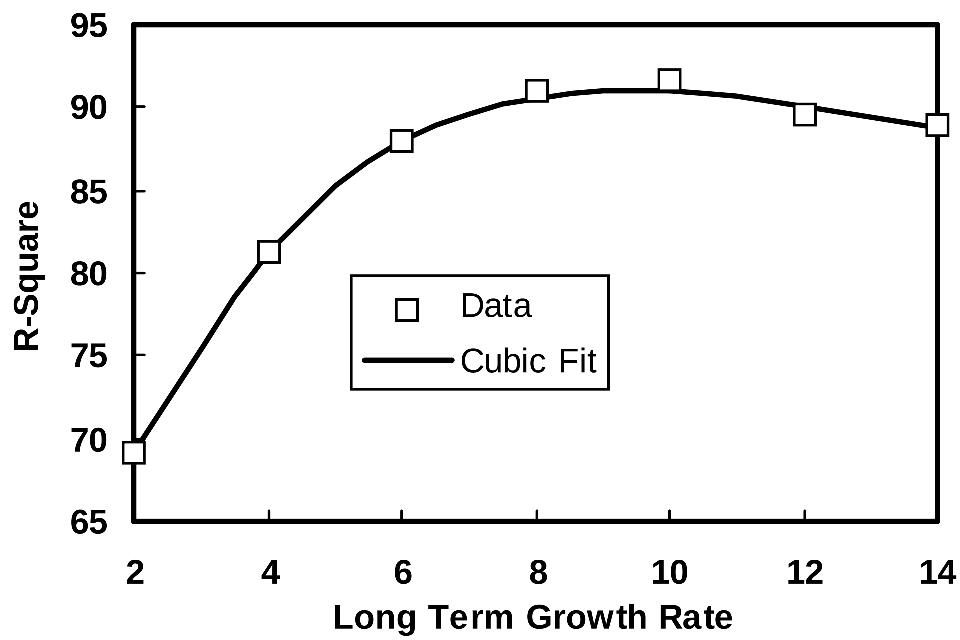


Fit by Geography of Industrial Products Against the Gompertz (General Sales Growth) Curve

Notice on the figures that all categories tend to have some extremely good fits. Most of the deviations category is due to a group of poor fitting members. As a set, the non-industrial products (pharmaceuticals, consumer, and farm products) have a poorer fit than the industrials. There are some industrial products, which have a poorer than average industrial fit (explosives and amino-salts). However, these involve small sample sizes with correspondingly large expected errors. With the exception of Japan, the results by geographic area appear to be the same. The averages and the ranges are extremely close. Japan appears to be the only outlier.

6.6.2.5.2 Long Term Growth

The model includes an asymptotic limit of growth. In order to test the effect on various limits, the model was reanalyzed varying long term growth rates between 2 and 14%. 100 of the longest range series were used for this analysis. The following figure shows the average R-Squares of the resulting fits. The maximum agreement can be seen between 8 and 10% with a severe decline at low long-term growth rates. The line on that figure is a "least squares" fit of a cubic relationship to the data. The maximum R-Square based on that fit corresponds to a 9.15% annual growth, a value used in all subsequent fits.



Optimum Long Range Growth Rate

6.6.2.6 Conclusions and Implications

The data analysis appears to support our hypotheses within its limits. We conclude therefore that:

- There appears to exist a common curve that describes the growth of industrial products;
- This curve is well approximated by a modified Gompertz model;
- An asymptotic limit of long term growth (between 8 - 10%) is supported by this analysis; and
- There appears to be an abrupt transition from product growth to maturity.

Some of the potential implications from these conclusions could have board scope both on a theoretical and a practical basis. In discussing these implications we acknowledge the need for further verification of the results with a larger sample and with brand data.

6.6.2.6.1 Common Sale Growth Curve is Universal

If there is a common sales growth pattern then either: (1) all successful industrial products have been produced and marketed effectively in the same manner or (2) changes in marketing strategy and mix for industrial products do not effect sales dynamics during the product growth period. If either of these are correct then traditional function and objectives of industrial marketing need to be reexamined.

Furthermore, if the sales growth pattern is constant then the maturation, which differs among products, is separate process. The conditions for maturation, therefore, do not depend on the dynamics of growth. The potential causes of maturation need to be reexamined.

If the model for sales growth is predictive it could provide a bound on sales forecasts during the growth phase. The common sales growth pattern would act as a upper bound for these products.

6.6.2.6.2 Gompertz Describes Sales Growth

The traditional sales forecasting tool for growth products is the Logistics curve. As previously mentioned this curve has been successful in describing technological substitution and sales dynamics. It is shown in the appendix that under some conditions the Logistics model approximates the Gompertz. If the Gompertz curve is the correct form, we may need to reevaluate projects based on the Logistics model.

Theoretical business analysis rests on identifying the key important variables. The key predictive variable for the Logistics model is penetration. For the Gompertz model it is growth rate. In fact, growth rate is a "state" variable in the Gompertz model. It is only a function of its own past history. If the Gompertz curve is the best description of sales growth, the possible mechanisms for growth may need to be reexamined.

6.6.2.6.3 A Limit to Long Term Growth

The growth curve and its long-term growth limit apply to a single product group. If there is a limit to long term growth it constitutes a natural boundary, then breaking through that barrier requires a new invention. The long term growth limit may act as a definition of meaningful break-through rather than a change in technology.

If there is a natural limit to long term growth, it can act as an upper bound to long term forecasts, 10 and 20 years into the future. It should be noted that this is only an upper bound. As some point the product will mature.

6.6.2.6.4 An Abrupt Transition to Product Maturity

If the transition to product maturity is abrupt than control tables can be constructed to determine the likelihood that a product has matured. These control tables could be constructed from analysis of extensive historical data. Analysis of our present database could be used as a starting point for the development of this type of tool. However, brand data would be necessary to make it functional on a product basis.

6.6.2.7 Relationship Development

6.6.2.7.1 Sales Growth Models

6.6.2.7.1.1 The Gompertz Curve

The Gompertz curve the simplest penetration curve, which is a function only of the fractional rate of growth and its history. This is illustrated by examining the differential form of the model:

GOMPERTZ CURVE

$$\text{Penetration} = w = A^{R^{(T-T_o)}}$$

$$\frac{d\ell n(w)}{dt} = \ell n(A) R^{(T-T_o)}\, \ell n(R)$$

$$\frac{d\ell n(w)}{dt} = -b\, \ell n(w)$$

where **w** is the penetration and **A, R, b** are constants. **dℓn(w)/dt** is the fraction rate of penetration which is negatively proportional to its integral. In other words, the fractional rate of penetration is a simple function of its past history. It is not a function of the level of penetration or the market size. In this sense the Gompertz curve with a fixed penetration rate, **b** is independent of nature of the product who sales it is describing.

6.6.2.7.1.2 Gompertz verses the Logistics Curve

If the Gompertz model is the correct model for describing sales growth, the use of the classic logistics curve may introduce enough error to preclude the identification of common behavior. The logistics model is similar enough to the Gompertz curve, however, to allow reasonable description of the data. In fact, we will show that the logistics model approximates the Gompertz curve around the saturation point. Under these conditions, the Gompertz and the logistics curves are identical. We show this by looking at the differential forms of the expressions. For the Gompertz curve, the derivative of the logarithm of penetration is proportional to the negative logarithm of the penetration equation [4.1.1-3]; while for the logistics curve, the same derivative is proportional to one minus the penetration, equation [4.1.2-3]:

LOGISTICS CURVE

$$\text{Penetration} = w = 1/(1 + \exp(-b[T-T_o])$$

$$\frac{d(w)}{dt} = bw(1-w)$$

$$\frac{d\ell n(w)}{dt} = b(1-w)$$

where **w** is the penetration (either the market share in the case of substitution or physical sales for sales growth); **$T-T_o$** is time compared to a characteristic process time (commercialization for the Gompertz or the inflection point for the logistics model).

The negative logarithm of the penetration from equation, **-Ln(w)**, and one minus the penetration from equation, **1-w**, are approximately equal around the value of one. The error between the two functions increases as we go to lower levels of penetration. However, even for levels of penetration less than 50% the error is less than 30%. Since most commercial data available to academicians is likely to be for high penetration ranges, it is not surprising that little difference would be noticed.

6.6.2.7.2 Linearization of the Sales Growth Model

In order to simplify the curve fitting and testing procedures the modified Gompertz curve can be transformed into a linear form. We start with the general form of the sales growth model:

$$U = P_o(1+i)^{(T-T_o)}\,(U_o/P_o)^{R^{(T-T_o)}}$$

where i is the long term growth rate. The other parameters have the same meaning as in previous equations. The long term growth rate, **i**, and the penetration parameter, **R**, are considered to be universal constants (**i** = 0.08, **R**=.0.77), while the other three parameters: **P_o**, **T_o**, and **U_o** are specific to the particular case being studied. Taking the natural logarithm of this relationship we obtain:

$$\ell n\{U\} = \ell n\{P_o\} + T\,\ell n\{1+i\} - T_o\,\ell n\{1+i\} + R^{T}\ell n(U_o/P_o\}/R^{T_o}$$

We now take transformed independent and dependent variables:

$$X = R^{T}$$

$$Y = \ell n\{U\} - T\,\ell n\{1+i\}$$

After substituting these variables results:

$$Y = \ell n\{P_o\} - T_o\,\ell n\{1+i\} + X\,\ell n\{U_o/P_o\}/R^{T_o}$$

This is a linear function with **$\ell n\{U_o/P_o\}/R^{T_o}$** being the slope, **B**, and **$\ell n\{P_o\} - T_o\,\ell n\{1+i\}$** being the intercept, **A**: This is therefore, a two parameter model. If the long term growth and the penetration parameter are considered specific to the particular product, as is done, traditionally, using the logistic model the model has four parameters.

6.6.2.7.3 Testing universality of the General Sales Growth Curve

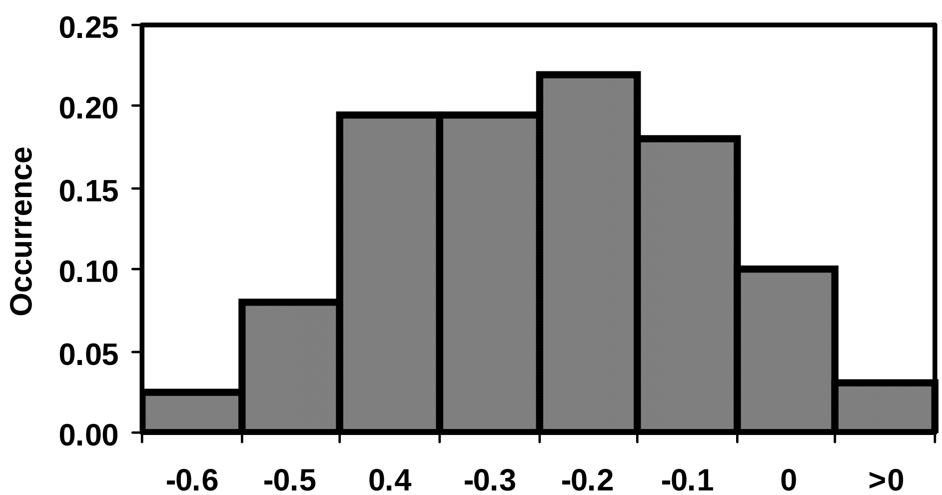
The model describing common sales growth behavior that have both universally fixed parameters and parameters specific to the application. For this case the potential and the starting date are specific to the product. The other parameters are assumed to be universal and therefore relatively constant among all series. In order to test agreement with the Logistics and Gompertz models least squares estimates of both types of parameters have to be made.

A multi-stage iterative process was used to fit these parameters using a data base of over a hundred series. The stages are:

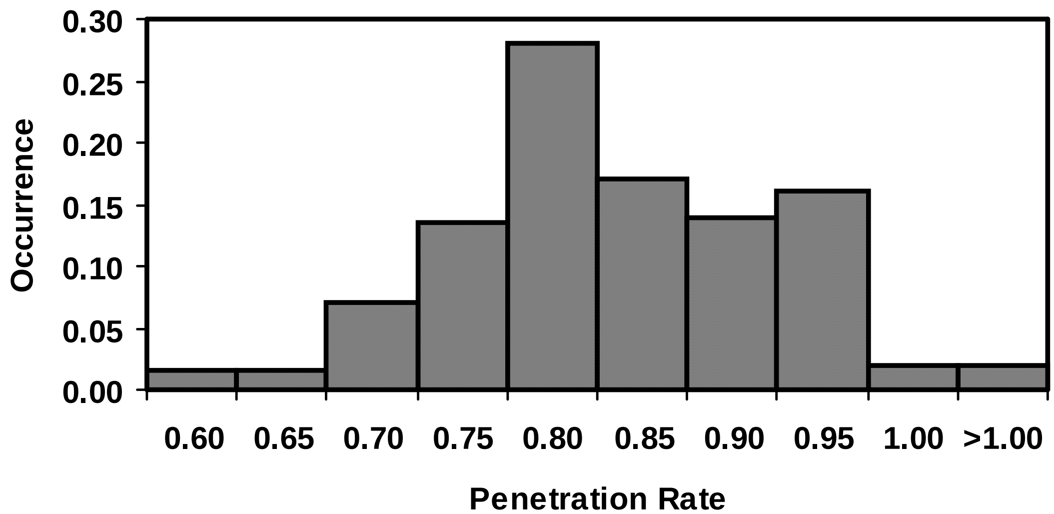
1. Initial estimates of the case specific parameters (potential and starting time) are obtained using assumed values of the universal parameters;
2. The "universal parameters" are estimated for each series separately based on the previous estimates of potential and starting time;
3. New estimates of the universal parameters are obtained based on the average from each series;
4. These new estimates are used to reestimate the case specific parameters in the next iteration.

The following figures show the resulting distributions of growth parameters for the Logistics and the Gompertz models respectively, based on a single iteration. It should be noted that the range of the Logistics parameter is much larger than that observed for the Gompertz curve with a 53% and only 10% coefficient of variation respectively.

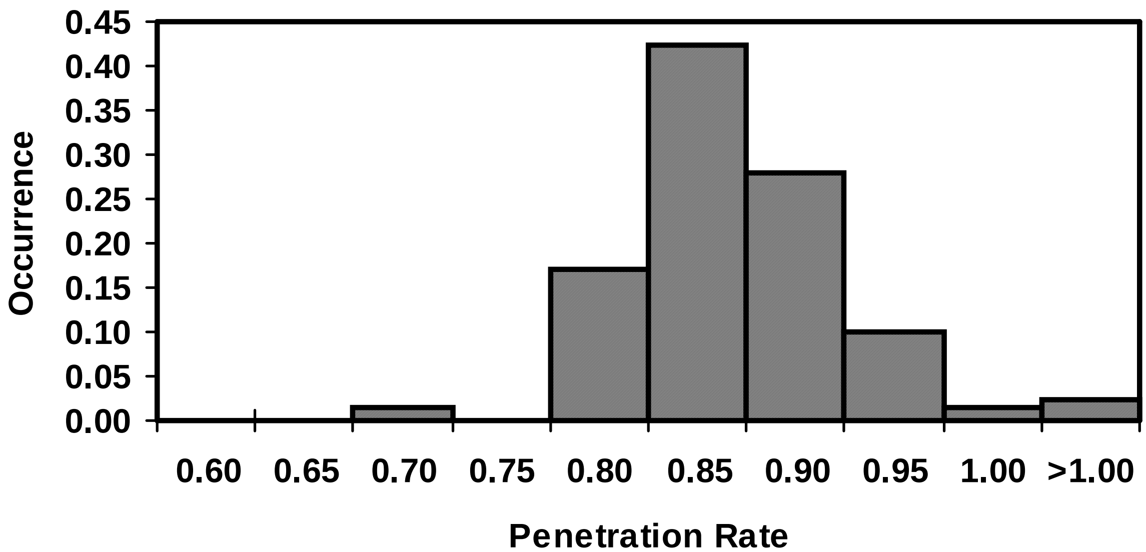
An iteration of the procedure could theoretically lead to improved estimates of both the potential and growth rate parameter. This requires two linear expressions, one for the potential and one for the growth rate parameter. Such expressions can be derived for the Gompertz curve. But unfortunately, not for the Logistics model. The following figures indicate the resulting distribution after 10 and 25 iterations. This reduces the coefficients of variation to 7 and 5.6% respectively.



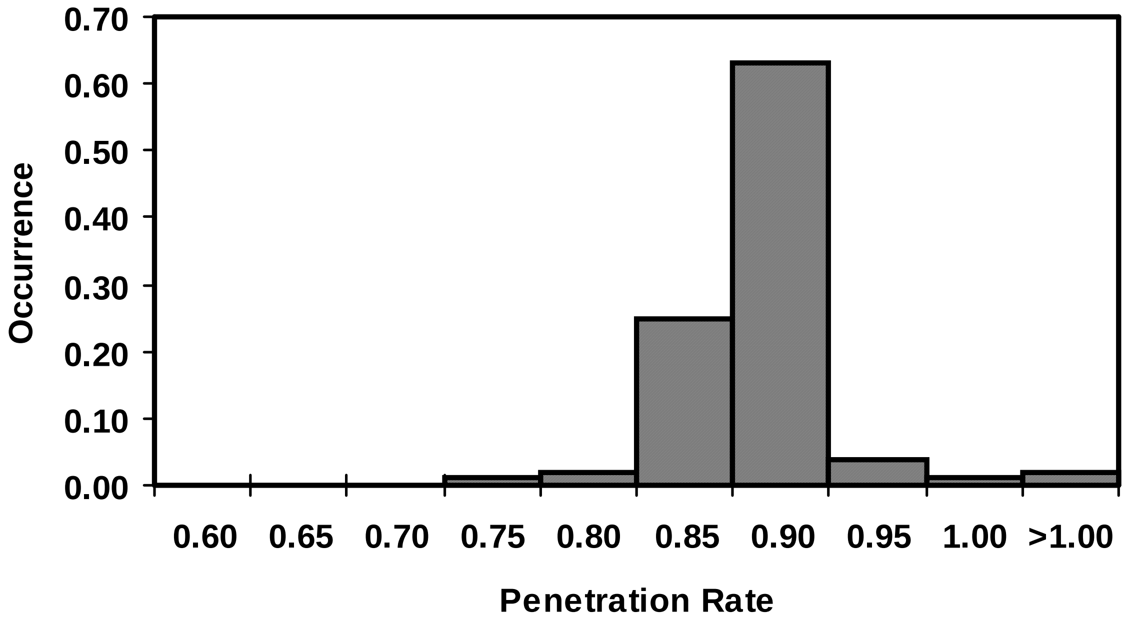
Distribution of Logistic Penetration Rates



Distribution of Gompertz Penetration Rates with No Iteration



Distribution of Gompertz Penetration Rates with 10 Iteration



Distribution of Gompertz Penetration Rates with 25 Iteration

6.6.2.7.4 Testing Abrupt Maturation

We observe that physical sales tend to break from growth curve abruptly. In testing the model in the previously mentioned procedures, we used a set of rules to exclude data points. We tested the effect of different rules on the subset of series that had long term data.

Rules for data exclusion -- due to the transition from growth to maturity -- should be based on the expected deviation between the growth curve and a mature product. Mature products tend to grow with their segment of the national economy. This may vary between 0 to 3%. Since the long term asymptotic growth was found to exceed 8%, this results in annual deviations of between 5 to 8%. We tested deviations below this level.

A number of points demonstrating consistent behavior are necessary to confirm a transition from a growth phase to maturity. Three data points were considered adequate for this determination. The fewer points used, the less reliable the decision rule is likely to be, however, the larger the number of points, the fewer sets of data can be used for analysis. The rules consist of:

- Consistent positive or negative deviations - Most deviations are negative, conforming to maturation. However, due to the aggregation of data, the series could consist of multiple products and showed multiple growth periods;
- The deviations from the curve are expected to increase linearly;
- The acceptance of early points follows our previously noted rule.

6.6.3 PRODUCT LIFE CHARACTERISTICS

A "Free Growth" Approach using the General Sales Growth Curve^SM

Events and mechanisms that effect sales dynamics are often difficult to identify or characterize. Traditional approaches of analysis consist of constructing models sufficiently complex to capture the phenomena. An alternative approach is to assume those products and technologies tend to undergo "Free Growth" over time periods. Deviations from that "free growth" can be identified and associated with underlying mechanisms. This paper discusses this approach using the General Sales Growth Curve^SM and analyzing a range of conditions and effects.

6.6.3.1 Introduction

The dynamics of technology and product sales growth can be very complex. War and recession can effect the sales levels at any time. All products eventually mature and most will eventually be substituted. Competition enters markets and radical new markets are found. All of these events make analysis and forecasting challenging.

Traditional, modeling approaches describes phenomena but does is of little help in identifying when or if they have taken place[55]. An alternative approach, however, does provide a means of identification and analysis. This approach is based on an assumption that the dynamics of technology substitution and sales are dominated by "Free Growth" into a specified potential. The complexity of sales dynamics is then interpreted as either local deviations or changes in the market potential.

This is in stark contrast with the traditional assumptions captured in the logistics type models. The general rate-form of these models are:

dU/dt = Growth Function X Maturation Function

The growth function is often taken as proportional to the physical sales, U. For the

[55] Bass, F.M., "A New Product Growth Model for Consumer Durables", *Management Science* 15 (1969), pp. 215-227

Chaffrey, J. M., Lilien (1980) *Market Planning for New Industrial Products*, New York, John Wiley

Horskey, D. "A Diffusion Model Incorporating Product Benefits, Price, Income and Information", *Marketing Science*, 9 (1990) pp. 342-365

Jain, D., Mahajan, V., Muller, E., " Innovation Diffusion in the Presence of Supply Restrictions", *Marketing Science*, 10 (1991) pp. 83-90

Mahajan, V., Muller, E., Bass, F. M., "New Product Diffusion Models in Marketing: A Review and Directions for Research", *Journal of Marketing*, 54 (1990), pp 1-26

Oren, S. S., Rothkapt, M. H., "A Market Dynamic Model for New Industrial Products and Its Applications", *Marketing Science*, 3 (1984) pp. 247-265

Schmittlein D., Mahajan, V. "Maximum Likelihood Estimation for an Innovation Diffusion Model for New Product Acceptance", *Marketing Science*, 1 (1982) pp. 57-78 classical logistics curve, the maturation function is taken as the difference between the present sales, U, and some saturation point, though many other forms have been suggested to cover other phenomena[56] [5,11,12,14]. In order to handle competing products and technologies, an array of rate expressions are used. This approach assumes an integrated view of phenomena is needed to adequately capture the effects[57] [6,7].

The "Free Growth" approach takes a local perspective. It assumes that there is a normative growth trajectory that sales and technologies follows if constrained by outside influences. Noted the conditions where the sales deviated from that normative behavior signals constraints. This approach is analogous to classical Newtonian particle mechanics, where objects once given an impulse will travel along a known trajectory unless acted upon by an outside force. Similarly, we assume that sales will follow a known trajectory into its potential unless that potential changes or an outside factor effects it.

6.6.3.2 General Sales Growth Curve[SM]

As previously noted, inherent in the logistics model, is the assumed exponential free growth of products. This predicts that unless limited by some constraint, sales should grow exponentially. While this is not a bad assumption for some matured products and technologies, it does not represent the growth phase of the technology and product cycle. Lieb as shown that an alternative two parameter model, referred to as the *General Sales Growth Curve*[SM], *GSGC*[SM], fits data remarkably well and significantly better than the exponential during growth [8]. Furthermore, the curve is an excellent predictor of sales compared to the exponential even using sparse data [9].

The General Sales Growth Curve describes sales trajectory starting with a high rate of growth which decrease continuously approaching an constant asymptotic limit. There are several functional forms of the curve that can be used, however, the modified Gompetz form is most convenient:

[56] Karmeshu, V.B. L., S. Karcher, "Model Innovation Diffusion with Distributed Time Lag", *Technological Forecasting and Social Change*, 34 (1988) pp. 103-113

Meade, N., "Technological Substitution: A Framework of Stochastic Models", *Technological Forecasting and Social Change*, 36 (1989) pp. 389-400

Raz, B. , I. Assa, "A Model of "Coupled" Technology Transfer: A Logistic Curve Approach", *Technological Forecasting and Social Change*, 33 (1988) pp. 251-265

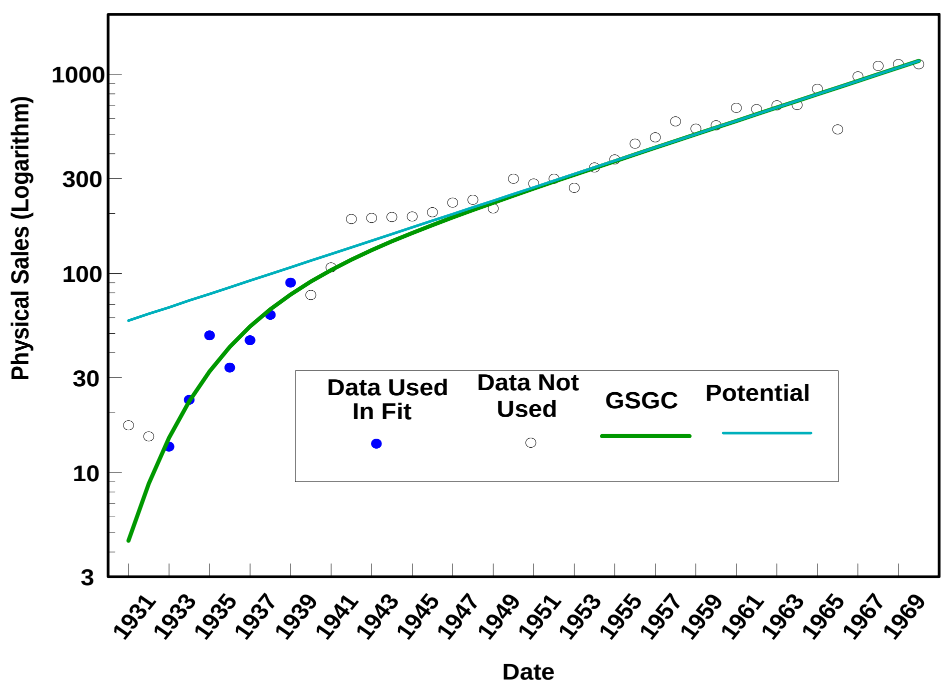
[57] Lakhani, H., "Empirical Implications of Mathematical Functions Used to Analyze Market Penetration of New Products", *Technological Forecasting and Social Change*, 15 (1979)

Lee, J. C., Lu, K. W., "Algorthm and Practice of Forecasting Technological Substitution with Data Based Transformed Models", *Technological Forecasting and Social Change*, 36 (1980) pp. 401 - 414

$$U = P_O(1+i)^{\{t-t_O\}} (U_O/P_O)^{R^{\{t-t_O\}}}$$

where **U** is the physical sales; **i** is the universal long term growth rate (0.08) and **R** is a universe parameter (0.77). $\mathbf{P_O}$ is the market potential in the year of commercialization, $\mathbf{t_O}$. $\mathbf{U_O}$ is the physical sales volume in that year. This relationship can be rearranged, resulting in a two- parameter model for curve fitting.

Soybean Product in the US



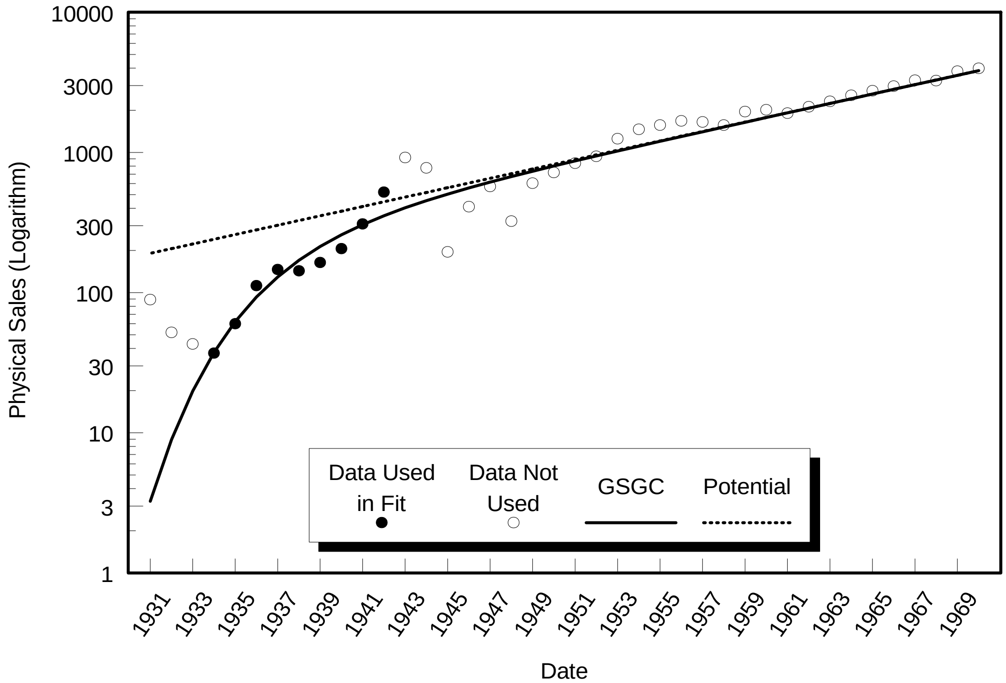
Parameters **A** and **B** are fit by regression. The figure above shows an example of the fit of data to the GSGC. Seven data points were used for the fit the curve, which extended over almost forty years of growth.

6.6.3.3 Morphologies of Growth

Depression & War

The curve fit covers only the region of growth. In case, some invention or use of soybeans took place during the 1930's that generated the growth process. Clearly the great depression did not affect this product. A positive deviation can be noted also during World War II.

A similar pattern is shown for aluminum production in the following figure.



The growth starting in the 1930 is likely due to the introduction of all metal (aluminum) aircraft during this period. The large positive deviation during World War II was likely associated with War production of aircraft as was the drop in sales shortly their after due to excess availability of scrap metal.

Increased production during war-time is not universal. In the case in the above figure for example, product decreased in that category during the war years. But here again, there was little effect of the depression[58].

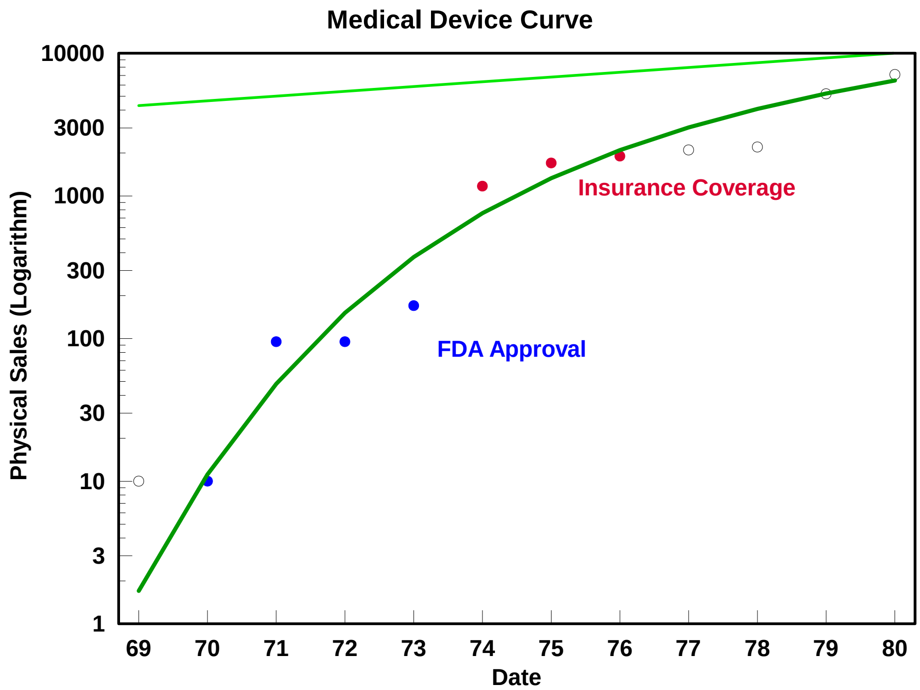
Government Regulations

In some cases, qualification and regulation affects the morphology of the sales growth. The following figure shows the growth of a new medical procedure. This shows typical breaks in the curve characteristics of a review and registration process[59].

---

[58] We suspect that economic recessions have more affect on pricing and revenue than on physical sales in most categories. This is is an area for future research.

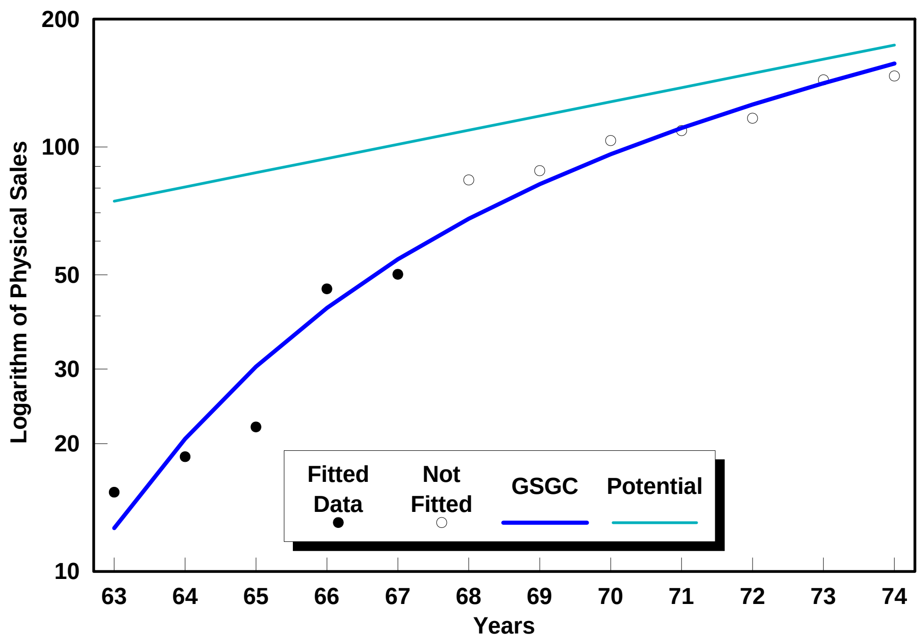
[59] Because of the broken nature of the growth curve of this type of product the corresponding measurements of goodness of fit are usually poorer than other cases.



Similar affects can be seen in other government influenced industries. In cases, where the market is mainly military advanced aerospace, sales will jump in response to new systems

Pricing & Capacity Limitation

The following figure shows a typical growth pattern.



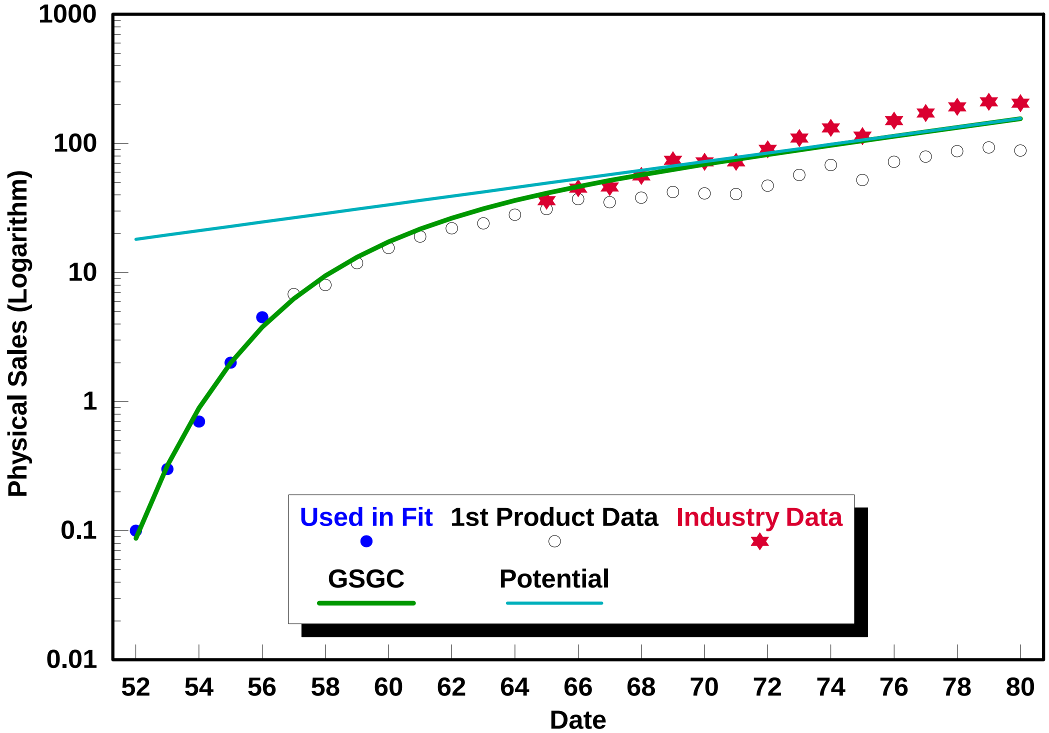
Note that during several periods sales drop below the historical curve but returned. These drops can often be actual reductions in sales for the year; but more often they represent growth that is below the historical trend. Unless these deviations are associated with the economic cycle or government actions we associate them with either capacity limitation or aggressive pricing.

The inability to reduce price or to grow value to the customer can result in a permanent deviation from the curve. This is the maturation process and is often associated with an inability to reduce cost. Since all products will eventually mature[60], one can assume that at some point, costs can not be sufficiently reduced to maintain a continuously decreasing effective price.

6.6.3.4 Competitive Entry

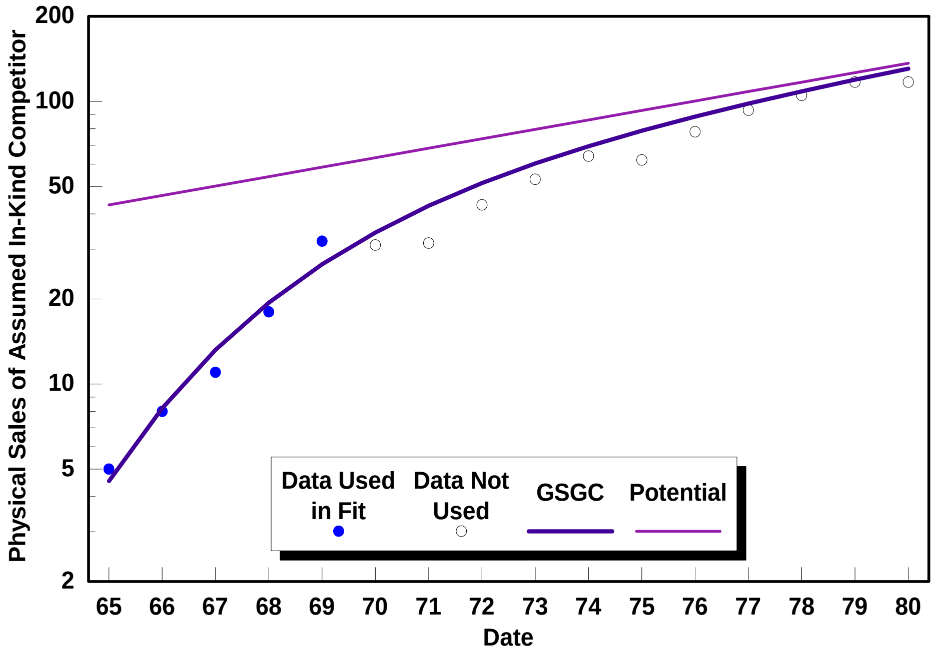
The previous phenomena did not influence the effective market share that the product or technology was growing into. This is not the case when in-kind competitors enter an market that had been previously single sourced. The following figure shows that situation.

[60] Since there is no upper limit to growth , without maturation, we would be hip high in everything.

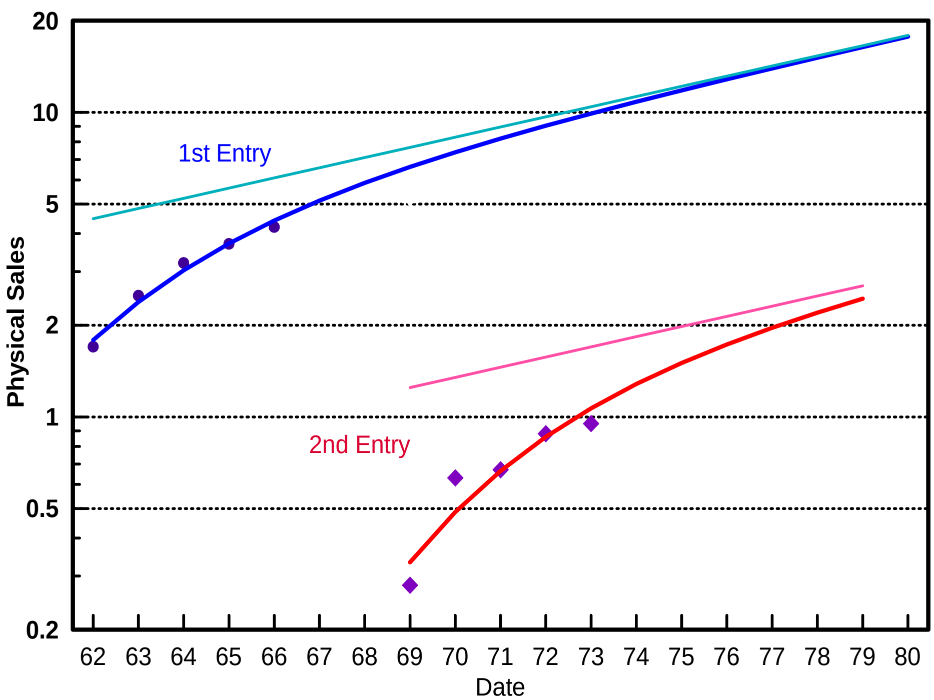


In this case other competitors have entered the market making the same product. The projection from the earliest five years of data continued to be a good forecast for the industry as a whole; but no longer tracks the original supplier.

We can estimate the sales of the other competitors as the difference between the industry and the sales to the original supplier. That data is shown on following figure. Here the GSGC also shows good agreement indicating that the sales from the other suppliers also undergo "Free Growth" into their own potential. This implies that the market potential with its associated market share is established early during the introduction of additional competitors.



The following figure shows another case of a second competitor entering the market. In this case actual data on the sales by the second competitor was known.

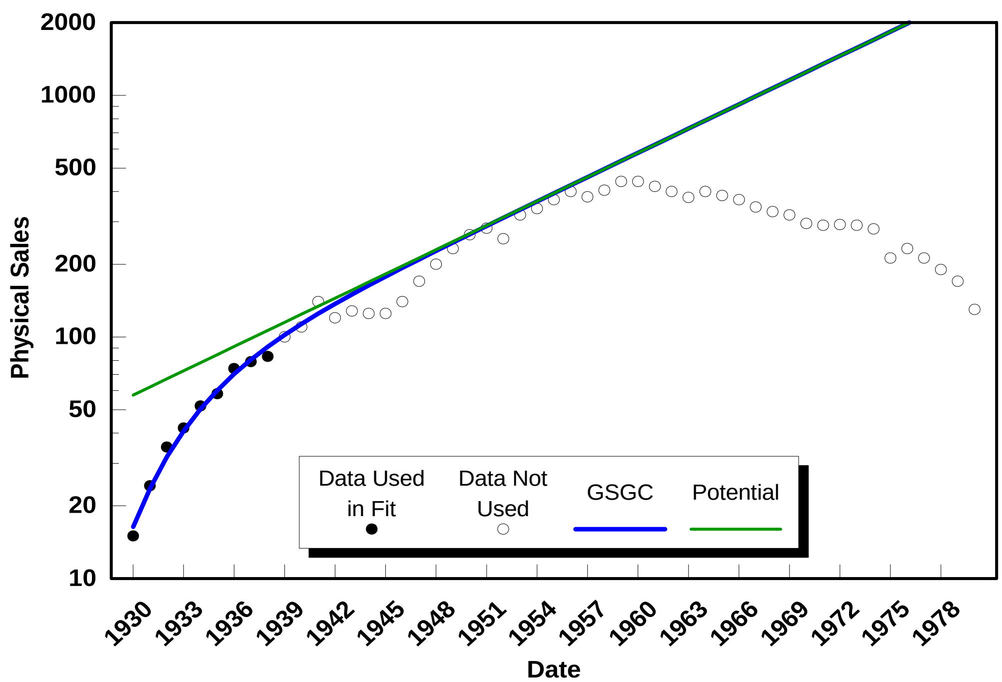


Note on this figure, that the entry of the second competitor had an immediate effect on the sales of the first entry.

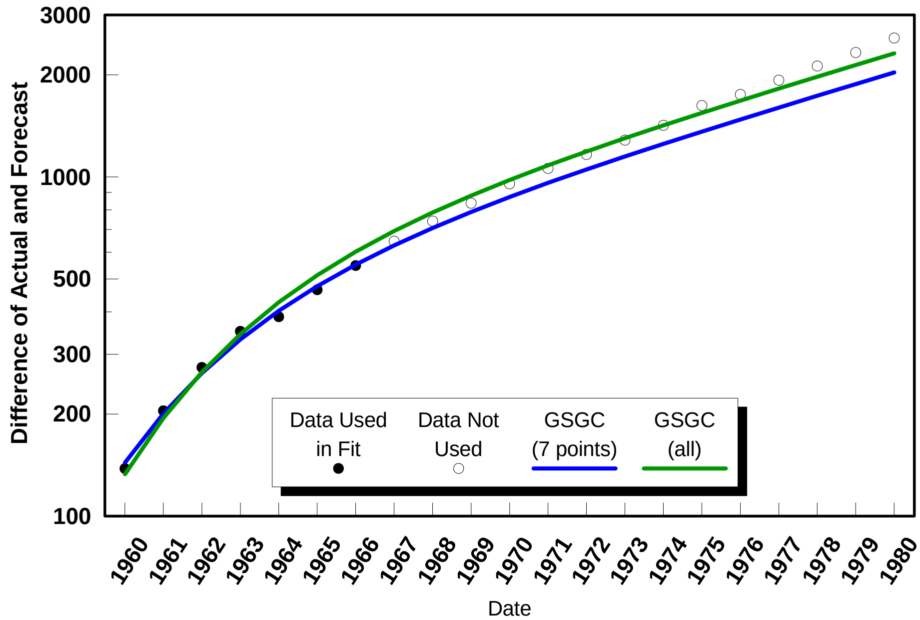
Estimating the future impact of a new competitor is often a prime concern to the established supplier. Relying on actual sales can greatly understate the competitive challenge. The eventual share can be estimated based on the share of the market potentials. In the case shown on the above figure, this measure is very close to the market share in the last year noted (1979).

6.6.3.5 Substitution

In-kind competition, discussed above, does not in itself result in the maturation of the industry but may affects the sales from each participant in the market. Functional competition, on the other hand, can not only mature the business but lead to its decline as is shown in the following figure.



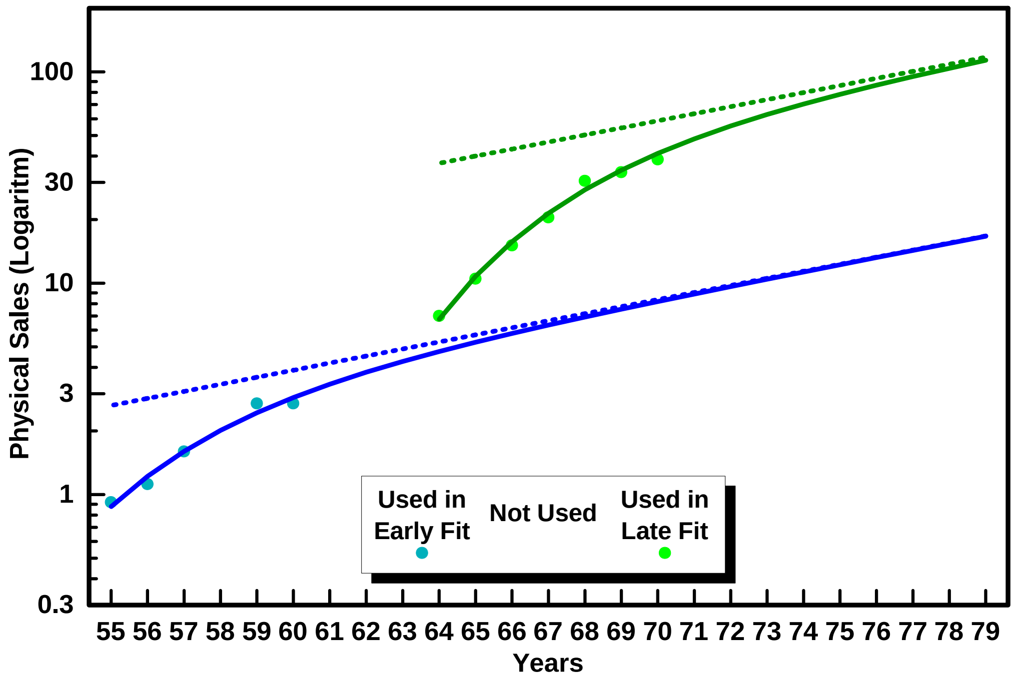
Here we have a case of an originally revolutionary product which created its own industry. It had 30 years of growth followed by over 20 years of decline. But what could cause such a decline. Most likely it would be one or more products that performed the same function but was significantly better or cheaper. If this is the case, one would expect that these new products would also follow the GSGC as shown on the following figure.



Notice that the curve uses only the early data and under estimates of sales. This is likely to be due to an increase in the market potential from applications previously unavailable with the older technology. However, this error is still surprising small.

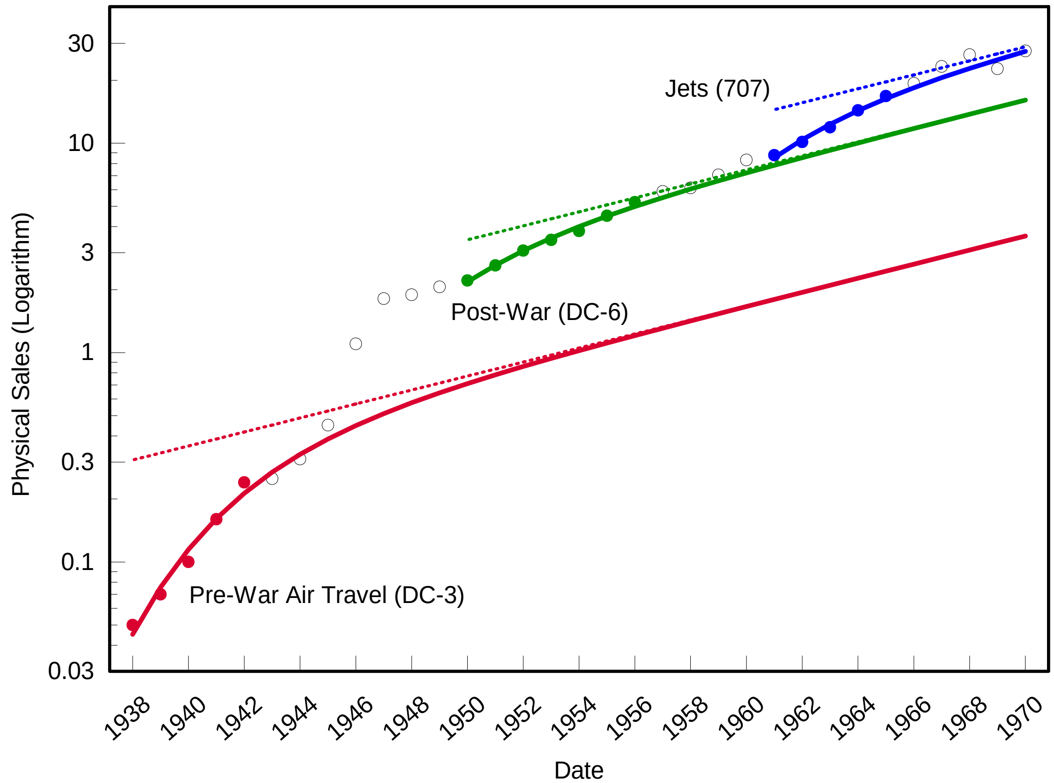
6.6.3.6 Renewal

Most of previously discussed phenomena showed a negative deviation from the GSGC. However, there are occasions where sales indicate the start of a new growth period. The following figure shows a large but not atypical product renewal.



The renewal process is characterized by an increase in the market potential. In this case, the market potential has apparently increased by a factor of 10.

The following figure shows international scheduled air travel with two renewal periods.



Such renewals are associated with some "invention" or change that materially affects how the product or technology can be used. It is an expansion of the applicable market. It should be noted that this invention may not be part of the product but reflect how the product is used. Furthermore, dramatic product and process improvements in the product may not affect the market potential. In the case of international air travel, these renewals can be associated with major changes in aircraft.

6.6.4 NEW PRODUCT SALES FORECASTS

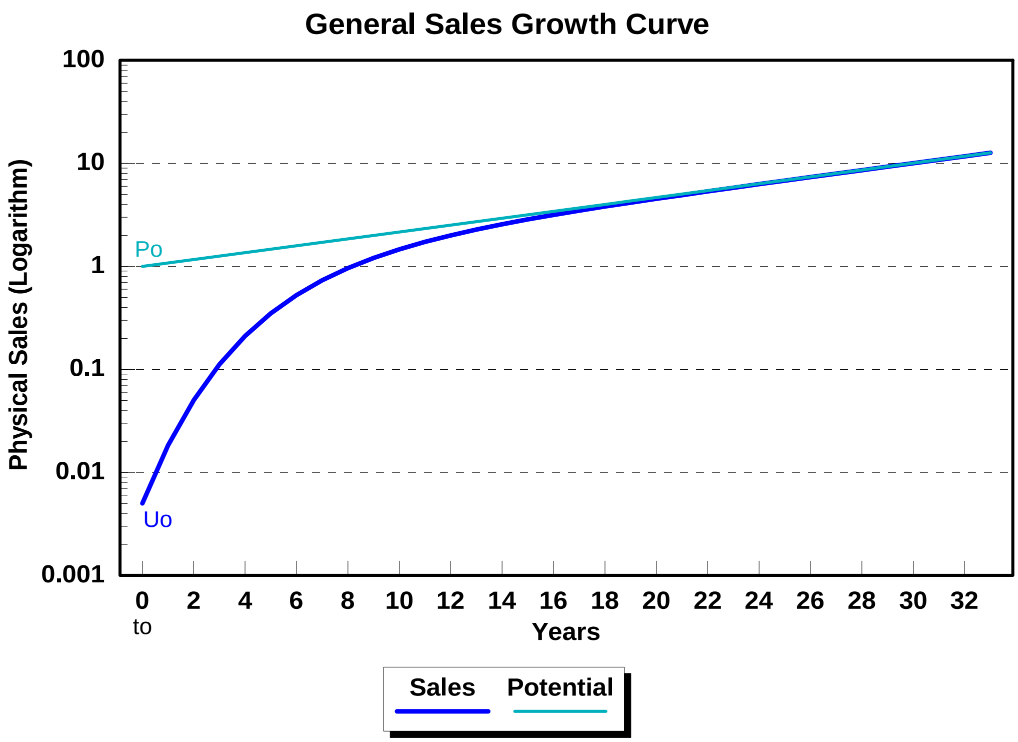
For new product forecasts, commercial data is usually not available making it infeasible to use the curve fitting techniques to the General Sales Growth Curve. However, estimates of the market potential, P_O, and the year of commercialization, t_O, can be estimated. Physical sales, **U,** over time, $t-t_O$, can then be estimated from the following relationship:

$$U = P_O(1.08)^{\{t-t_O\}}\,(U_O/P_O)^{0.77^{\{t-t_O\}}}$$

if the initial physical sales, U_O can be estimated. For new business forecasting initial physical sales is usually not known. However, we often assume that $U_O/P_O = 0.005$. While this is not a universally established value, without additional information it appears to be useable. If a better estimate is available it should be used. The resulting relationship is:

$$U = P_O(1.08)^{\{t-t_O\}}\,(0.005)^{0.77^{\{t-t_O\}}}$$

Graphically this result is shown below. This is then used as a first estimate of the growth of new product sales. It should be noted, that the market potential must be adjusted for market share. This is often done using the Broken Stick Rule assuming an eventual market position.


General Sales Growth Curve
Physical Sales (Logarithm)
100
10
1
0.1
0.01
0.001
Po
Uo
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
to
Years
Sales
Potential

6.6.5 PRICE/VOLUME CURVES

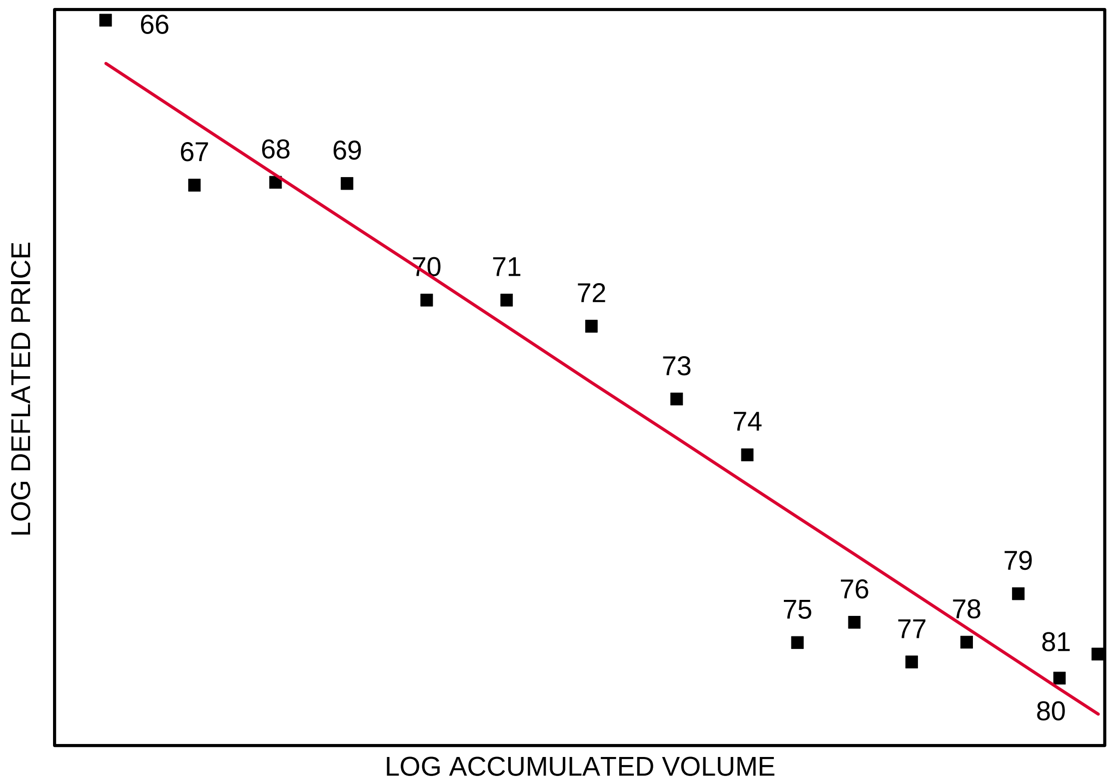
Forecasting prices is difficult at best but often critical. There are usually two issues: (1) what are reasonable prices for new products and (2) how will competition force prices down. Both problems are influenced by the relationship between price and volume. The first problem can be approached using value analysis and limited by the "Exclusion" curve. This concept is an extension of the ideas behind economies of scale.

The second problem is more difficult and based on assumed relationships between price and cost with production volume. There are two mechanisms that have been proposed for driving costs down: (1) Learning, and (2) Economies of Scale.

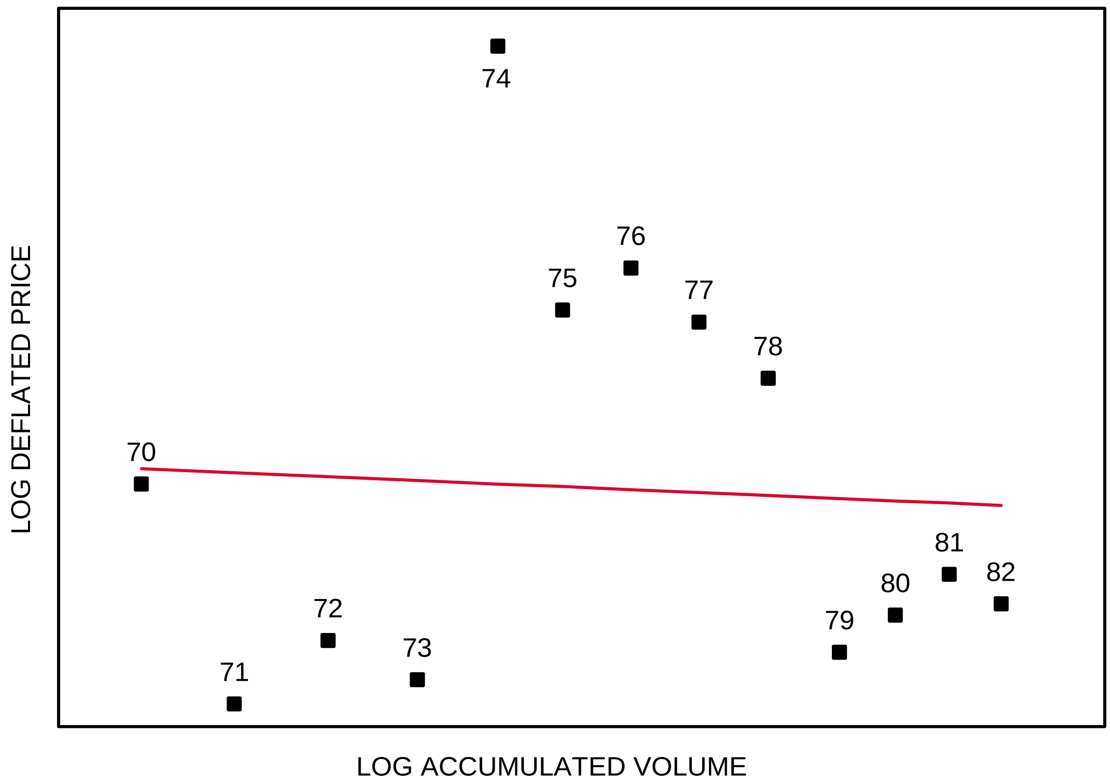
6.6.5.1 "Learning" and "Experience" Curves

The learning curve approach assumes that price should decline with the total production. This is based on the premise that experience in production should lead to methods that reduces cost and therefore, allowing for a decline in price. Price reduction is then driven by a desire to increase volume and share[61]. The basic approach is to plot the deflated or real price (adjusted for inflation) against the accumulated production or sales. As shown below, most capital intensive products show a decline in price. This curve is for polycarbonate plastics.

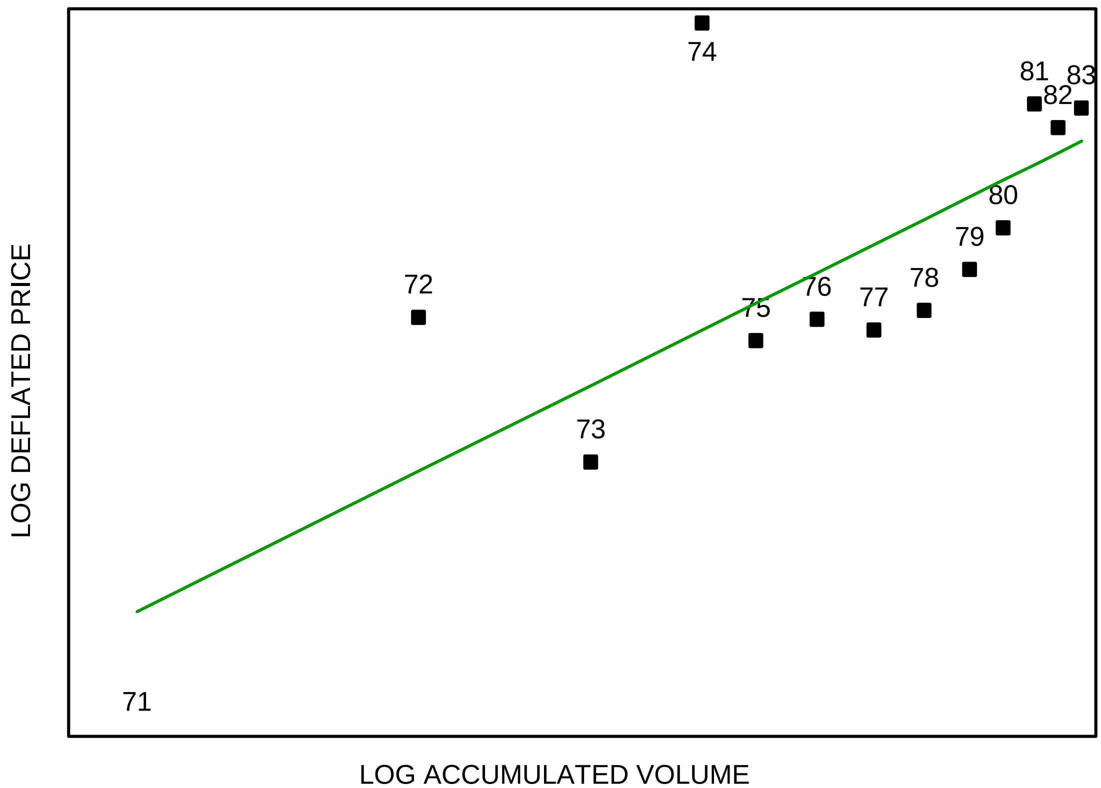
[61] The Boston Consulting Group, (BCG) had sponsored this approach with the argument that PIMS data had indicated that market share is the key driver for profitability. By preemptively reducing price and buying share, a firm would capture the "high ground" and assure itself high profitability. Unfortunately, firms (Texas Instruments) who had tried this tactic had an extremely poor result.



The Boston Consulting Group has suggested a 30% reduction with a doubling of accumulated production. However, as can be seen in the two graphs below, this is not necessarily the case. For, high density polyethylene variations in feedstock costs greatly effect price, indicating almost a random change.



In the case of saran resin, the deflated price has increased with accumulated volume. This appears to be partially due to feed stock prices but more likely to high capacity utilization, limited competition and possibly reduction in actual capacity.

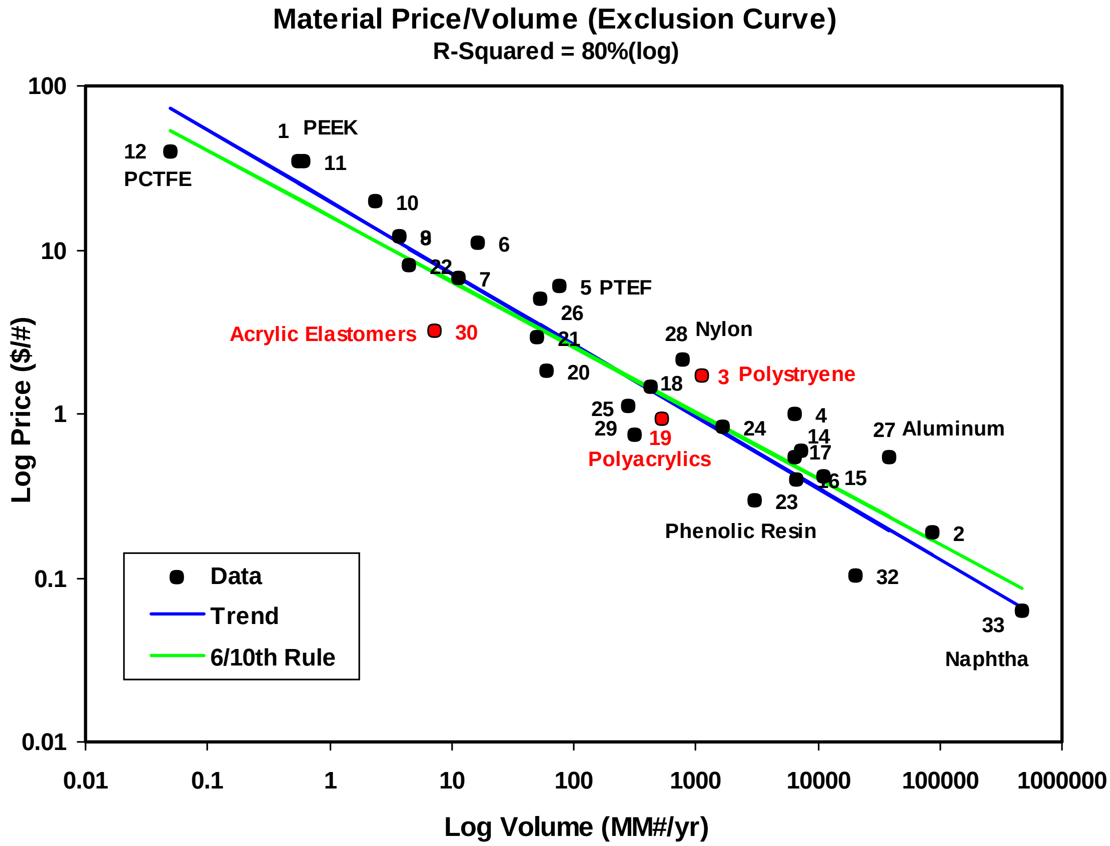


The general conclusion is that learning curve approach, while useful is not necessarily a good forecaster of price. The expectation of decreased cost with accumulated production needs to be based on some type of mechanism. In the case of assembled devices, such as in electronics, parts-consolidation is likely to lead to significant cost reduction and corresponding price reduction under competition. In cases, with high dependence on feed stocks or services, such large cost reductions are unlikely.

6.6.5.2 Price/Volume Analysis, "The Exclusion Curve"

From an economics viewpoint, costs may be highly dependent on "scale." The larger the plant capacity or business operations are, the less expensive per unit it is to produce products. Under competitive pressure, one would, therefore, expect price also to decline. This relationship is evident in the description of price versus shipment of materials, as shown below. This relationship is referred to as the "Exclusion Curve" in that it seems to exclude certain new material business situations, specifically, a very large volume demand for a high priced material[62].

[62] The actual originator of this concept is lost. At Dupont, it was well known by the early 1970's and probably goes back at least 10 years before. It was also known at GE in the early 1980's.



Notice that the data covers many orders of magnitude in both volume and price[63]. While the scatter of the data appears small in this plot and the correlation is very high, that variation corresponds to a broad range (almost of a factor of 2).

6.6.5.2.1 Time Stability

The regression fit of logarithm of price to the logarithm of volume corresponds to a "Power-Law" relationship:

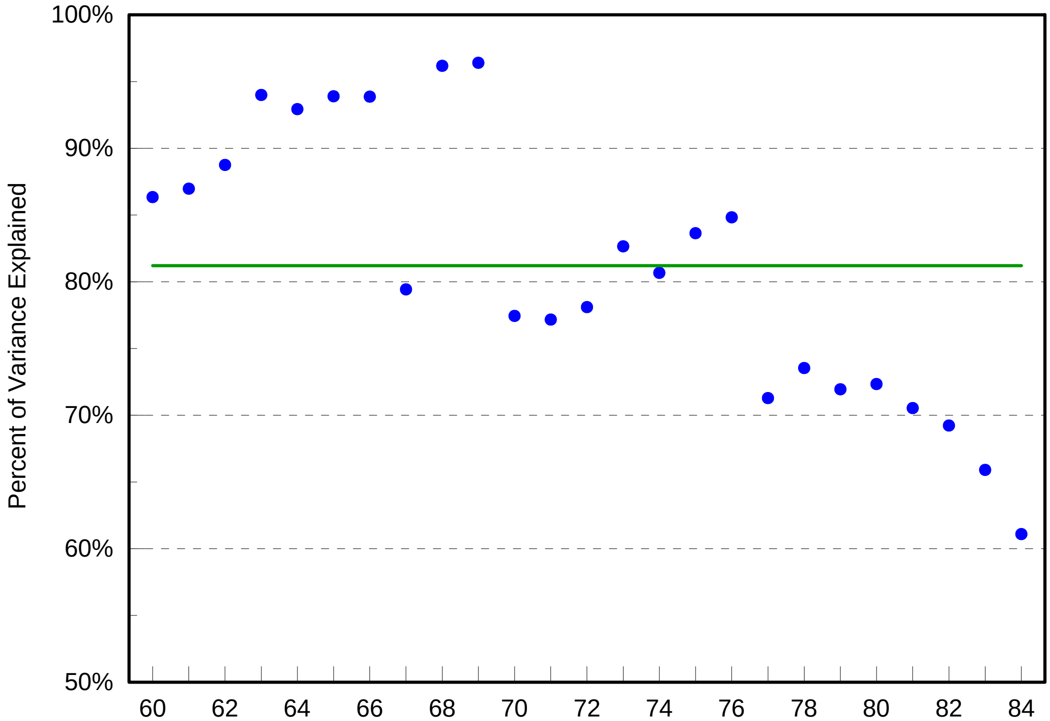
$$\textbf{Price} = \textbf{A} \cdot \textbf{Volume}^{\textbf{b}}$$

Below is the R-Squares for this regression fit over a 26 year period[64]. It should be noted that the number of points change over this period of time. The average R-Square is over 80%. This roughly means that over 90% of the variation in price is explained by scale

[63] The data in this curve are mainly organic or light weight (low density) materials. With more diverse density materials it is useful to express price and shipments in physical volume such as $/cubic inch, or millions of cubic feet sold. This puts the materials on a common basis. Even on this plot, the denser resins (fluoropolymers) show a corresponding higher price than would been the case using $/cubic inch basis.

[64] Earlier work has indicated that the general exclusion curve relationship has been stable since 1947.

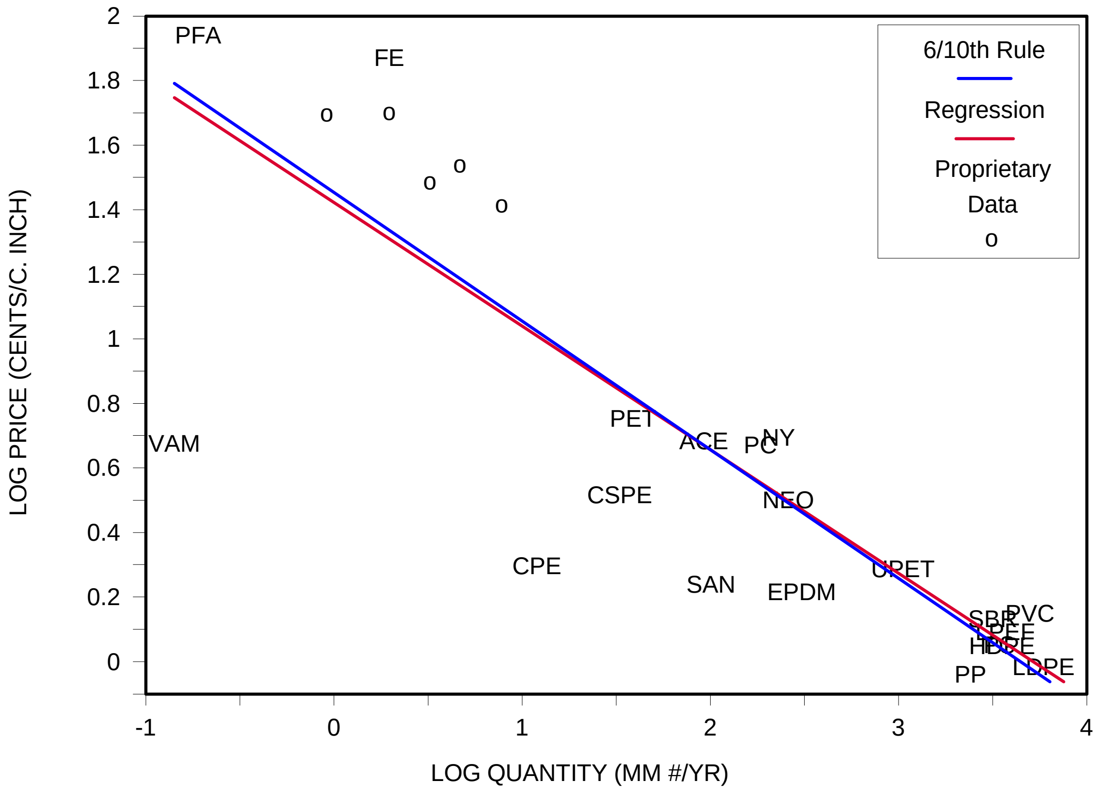
alone.



6.6.5.2.2 The 6/10th's Rule and "Power Law"

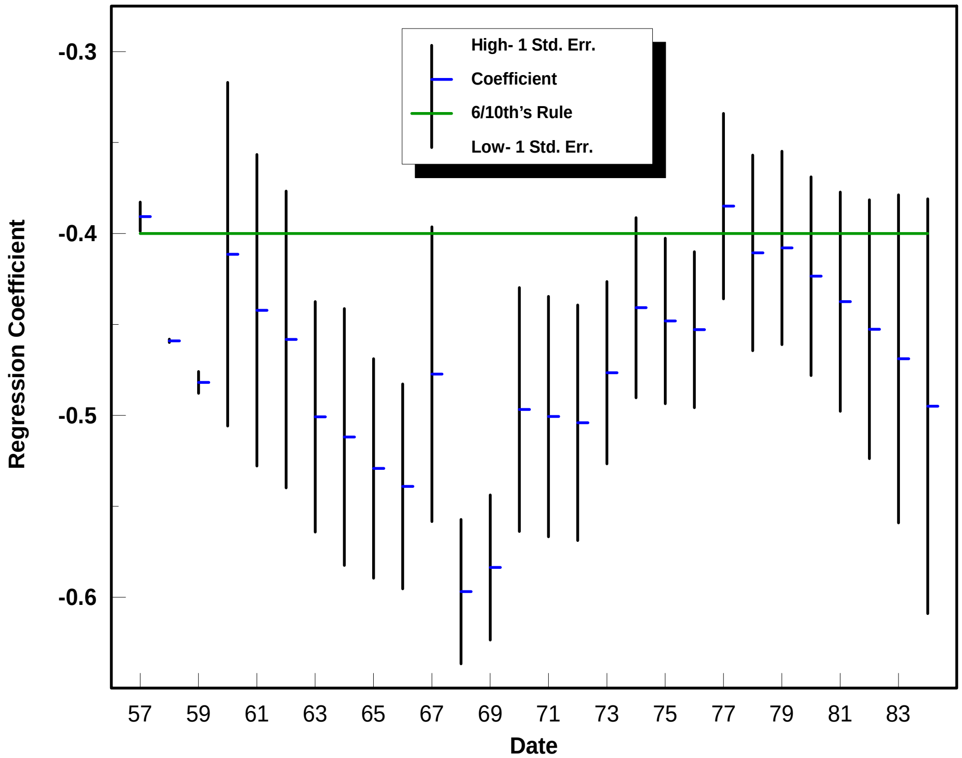
While a "power-law" regression curve gives the "best" fit to the data, there is motivation suggesting a universal relation based on a "6/10th" rule. The "6/10th" rule is a general chemical engineering cost scaling tool. This rule indicates that in general total capital costs of a chemical manufacturing facility increases with scale raised to the 6/10th power. This is a significant reduction in the unit costs with scale. On a unit cost or unit price basis this would correspond to a slope on the log price versus log volume[65] of -0.4. The regression curve is compared to the 6/10th rule on the chart below.

[65] This is the inverse of 6/10th (1-0.6)



1977 Polymer Resin Data

The fit for 1977 data is very good. However, it is not the case for all of the data available as shown on the chart below. The regression values may vary from -.38 to less than -.6.

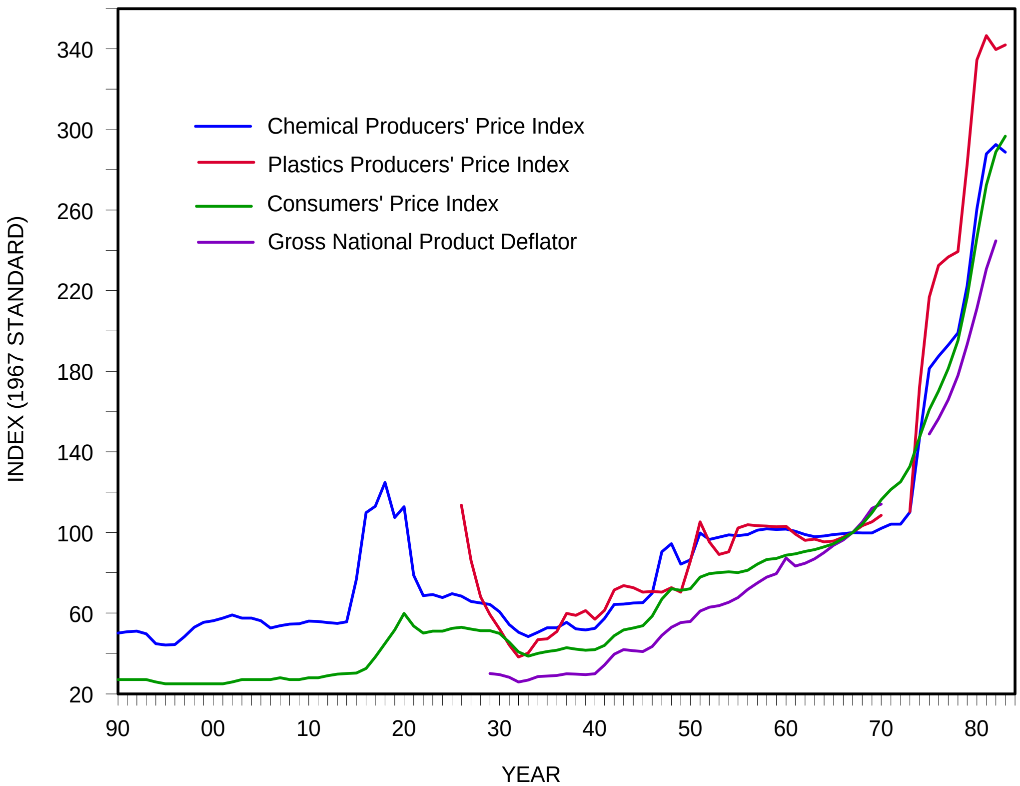


However, the error captured by the "6/10th rule" is significantly better than the average regression values as shown below.

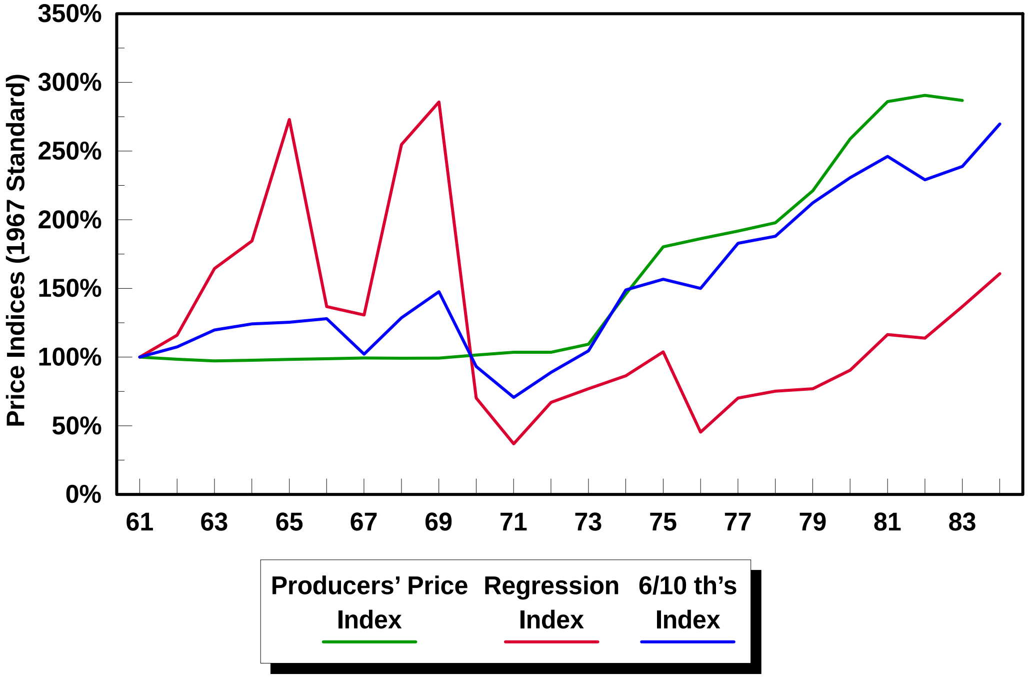
	Regression	6/10th's Rule
Average Absolute Error	65%	16%
Average Error	8.4%	1.4%
Variance Explained	19.3%	83.8%

6.6.5.2.3 Economic Indices

The preference for using the 6/10th rule is even more apparent when the annual data is standardized using price indices. The price index is a means of deflating prices to a common scale. For resins and other chemical materials there are generally four indices of interest as shown below.



Of these, the most general is the Producers' Price Index (for materials) since it covers all materials produced. The scaled coefficient for regression, A, for the power law model and the corresponding coefficient for the 6/10th rule model are compared with the Producers' Price Index in the chart below. Notice a good agreement between the index and the 6/10th's rule. This strongly supports the use of the 6/10th rule for price scaling.



Price Indices (1967 Standard)
350%
300%
250%
200%
150%
100%
50%
0%
61
63
65
67
69
71
73
75
77
79
81
83
Producers' Price Index
Regression Index
6/10 th's Index