An Examination of New Product Diffusion Models

By

Jean-Pierre Van de Capelle, Ph.D.

MBA Student, Rochester Institute of Technology

Xerox Corporation

A Research Monograph of the Printing Industry Center at RIT

October 2004

No. PICRM-2004-02

R·I·T Printing Industry Center
An Alfred P. Sloan Foundation Center

An Examination of New Product Diffusion Models

By
Jean-Pierre Van de Capelle, Ph.D.
MBA Student, Rochester Institute of Technology

Xerox Corporation

Printing Industry Center
An Alfred P. Sloan Foundation Center

A Research Monograph of the
Printing Industry Center at RIT
Rochester, NY
October 2004

With Thanks

The research agenda of the Printing Industry Center at RIT and the publication of research findings are supported by the following organizations:

Rochester Institute of Technology R·I·T 1829

Alfred P. Sloan Foundation 1934

Adobe

creo

RR Donnelley

Heidelberg

hp invent

IBM

MeadWestvaco

NexPress A Kodak Company

NPES

Standard Register

vigc Vlaams Innovatiecentrum voor Grafische Communicatie vzw Flemish Innovationcenter for Graphic Communication

Weyerhaeuser

The Document Company XEROX

Table of Contents

Executive Summary

This paper is based on the book, *New Product Diffusion Models* (Mahajan, Muller & Wind, 2000), and is organized as follows:

- The summary presents a short overview of the book, which addresses many issues of new product diffusion as well as some models that describe such diffusion processes. Among them are the basic Bass diffusion and multi-state flow models.

- The "Interpretations" section provides a in-depth discussion illustrating the author's view of multi-state flow models and their relation to new product diffusion models. This section goes beyond the material covered in the book and is intended to provide the reader with a better understanding of the Bass model and its underlying assumptions. This section also includes a discussion on parameter extraction techniques, which is continued in subsequent sections.

- The application of the Bass diffusion model for the Xerox DocuTech family of products is discussed.

- The application of the Bass diffusion model in the digital color press market is also discussed.

The outcome of this investigation shows that the Bass model requires from 6 to 10 years of sales to be a valid sales model, and that the model functions better if the market size is known. The examples show however, that once a few years of data are available and the market size is known, the model can predict sales quite accurately for a long time into the future, with very good precision.

For the Xerox DocuTech family, an accurate prediction of sales for seven years into the future, with annual forecast errors of 5% to 10%, could be obtained by using only the first six years of sales data. This is a remarkable accomplishment.

Using this model it is estimated that by 2010 the primary market for black-and-white digital presses will be fully penetrated. This means that an additional 8,000 to 9,000 units will be placed. As this does not include replacement purchases, the estimated remaining total opportunity for black-and-white DocuTech devices is well above 8,000 to 9,000 units for the next seven years.

According to the available data, digital color presses diffuse faster in the marketplace than their black-and-white counterparts. The total market size for digital color presses with a production speed of 100 ppm or less is estimated to be at least 70,000 units and it could be much higher than that. In any case, for market sizes up to 120,000 units, it was found that most of the sales of digital color presses will have happened by 2010 and that an additional 20,000 to 70,000 units will be sold by that time.

A word of caution is required in terms of the numbers: They depend heavily on the estimated market sizes for these products. It is felt that recent economic changes in some parts of the world may provide an even greater opportunity than that represented by the numbers found through this analysis. In other words, the estimated market sizes were based on sales data in existing markets. Including less exploited markets would have a positive effect on the estimated market sizes.

Summary of the Book

New Product Diffusion Models is an aggregate of chapters written by different authors on the topic of new product diffusion models, extending the original model of Frank Bass (1969).

It covers some possible applications of diffusion models, which, according to the editors, include the following:

- The timing of successive product generations
- The estimation of pirate sales
- The estimation of lost sales and market expansion
- Assessment of market saturation
- Determination of market value of companies
- Capacity decisions and operations planning.

Clearly, in order for such models to live up to such expectations their validity must be proven and understood.

The most basic diffusion model can be described as "single market, aggregate, non-explanatory, time-independent, firm action independent product diffusion". It is the basis for the simplest models originally proposed by Bass, and its validity will be examined in this paper by applying it to two examples.

The book discusses the relationship between firm strategy and product diffusion. It also discusses digital media as a means of communication and their influence on product diffusion.

It is argued in the book that diffusion models can be extended to multi-market and global product diffusion and that new product diffusion can be different in different marketplaces around the world.

Chapter 5 of the book attempts to model the influence of the marketing mix on product diffusion. The complexity of such models makes them unappealing from a managerial point of view.

The book gives an overview of a few more elaborate product diffusion models, including replacement and multiple purchases and/or competition. Some of the models are discrete in nature, while others are continuous in nature.

Chapter 9 discusses disaggregate, individual diffusion models, based on Bayesian probabilities, sometimes also referred to as hazard-rate or multi-state flow models. These models can serve as bridges between the continuous diffusion models and the probability models. This paper will describe the author's view on this topic and will delve into it more deeply. In order to help the reader understand the basic Bass diffusion model and its underlying assumptions.

The diffusion of some products can depend heavily on the existence and diffusion of other products. Chapter 7 of the book discusses diffusion models with multi-product interactions, such as substitutes, complements, and independent or competitive products. Chapter 7 also discusses the influences of marketing decisions on product diffusion.

Finally, the book gives an overview of different parameter estimation techniques for aggregate diffusion models without going into the details of how this can be done. Evaluation of the different parameter estimation techniques would require an in-depth literature study. In this paper a least-squares technique will be applied to two examples, and the fundamental problem of all parameter estimation techniques, namely the existence of local minima, will be explained.

Interpretations

This section of the paper contains the author's interpretations of the book used as the basis for this study. The author has developed the formulas and derivations in this section, unless otherwise stated. Since the book under discussion is a compilation of chapters by different authors, there is some lack of coherent explanations. As a consequence, it is not easy to understand the different assumptions made in the basic Bass model and its relation to the complicated market processes happening in the real world. This section briefly describes some of the marketing processes and their relation to the Bass model in an attempt to clarify the strengths and weaknesses of the model.

One of the basic assumptions of the Bass model is the time independence of the diffusion parameters. It is asserted that this is very much dependent on the market segment and on how this segment is defined.

Multi-state flow models offer a comprehensible overview of market complexity can be a bridge toward the Bass model. The book does not provide a clear description of these models. This section of the paper offers the author's view on such models.

MARKET SEGMENTATION

Markets can be divided into mutually exclusive, interdependent segments, and product diffusion can be studied in each of the segments, taking into account possible interactions between the segments.

In general, market segments are time-dependent in many ways. Each market segment can have many time dependent properties, which can be categorized as follows:

1. **Demographics**
   The constitution and size of parameters such as ethnicity, cultural background, age, and disposable income typically vary rather slowly over time.

2. **Psychographics**
   The constitution and size of parameters such as market members tastes; needs; emotions; educational; social and religious backgrounds; awareness; and knowledge can vary either slowly or quickly, depending on events in the marketplace.

3. **Market structure and context**
   Competitive, complementary, substitute, and independent products, with their features, functions, costs, and benefits can vary abruptly in most marketplaces. The supply chain and market channel structure, effectiveness and efficiency, cost of labor, and access to capital usually vary more slowly.

   Market communication channels (such as advertisement, Internet, and word of mouth) and regulations also can change quickly.

   Product is typically not abundantly available at the time of a product launch. Companies may regulate the availability of product in order to acquire customer feedback and still improve on the product features. This introduces a time dependence in the number of units installed in the field, that does not follow the diffusion process.

4. **Macro-economic influences on the market segment**
In most cases, macro-economic parameters, such as cost of money, GDP, price indicators (CPI, PPI), confidence indices, political systems, non-market-specific regulations, and global interaction (exchange rates, regulations, treaties, trade barriers, peace or war situations), vary rather slowly over time and can be anticipated to some degree. They may have a small or large effect on the market segment that is under investigation.

All of these parameters positively or negatively an influence on product diffusion and they usually vary over time, either quasi-continuously or abruptly. Recent economic changes in parts of the world may not yet have had the chance to affect some of the data used in this study. Conclusions based on this data may therefore have to be adjusted to reflect these macro-economic influences.

Time Dependence and Choice of Market Segments

The time dependence of the market segment is a function of the size of the segment. In general, the narrower the market segment the faster the segment will change. Simple diffusion models assume that diffusion parameters are constant over time. In order for this assumption to be valid, market segmentation needs to support it.

MARKET GROWTH OR DECLINE VERSUS SALES GROWTH

A change in market size is often described by a growth rate factor *g*:

$$m(t) = m(0) * (1 + g)^t \tag{1}$$

Here $m(t)$ is the market size at time t. Market size growth must not be confused with annual sales growth, which is caused by diffusion into a marketplace, even if size is constant. Market size changes reflect population size changes within the market segment. The market size for a company can change if the company decides to change its covered territory, for instance, by expanding into multiple cities or states or going internationally. In such a case the market size would change abruptly.

In the printing industry, for example, if the number of printers were to grow, then one could assert that the market size for presses would grow. Underlying this growth, however, is a more fundamental structure reflecting delivered goods and services, which may depend solely on the total population size of the country. This would result in a constant market size.

MULTI-STATE TRANSITION OR FLOW MODELS

The next paragraphs describe the author's views on the Bass equation and multi-state flow models and ways in which these models can be used to understand the aggregate and disaggregate diffusion models and the relationships between them. An extensive literature study would be needed to comprehend all of the publicly available material and this is beyond the scope of this study. Instead this section of the paper is meant to provide a theoretical and conceptual background to help the reader to understand multi-state flow models, the Bass diffusion equation, and the relationships between them.

Defining multiple states in a market segment depends on the product type. For instance for consumable products, the consumption per time period (week, month, year) is the relevant parameter, whereas, for durable goods, it is the number of items owned.

Since this study focuses on asset-type goods (e.g., durable goods) the customer states could be defined as show in Table 1.

s0	Has never adopted
s1	Had product before (but not anymore)
s2	Has one from primary market
s3	Has one from secondary market
s4	Has two products from primary market
s5	Has one product from primary market, one from secondary market
s6	Has two products from secondary market

Table 1. Definition of Customer States

The customer states are indicated as *s*0, *s*1, and so on. Clearly, this list is only a subset of all possible states in the market segment at hand. At any given time, each customer within the market segment is considered to be in one of the states. The term *primary market* refers to the market where the sale of the durable goods is made by the manufacturer to the first buyer of the goods. *Secondary market* refers to the market where customers sell goods to one another without the producer's intervention.

It is clear that a customer can change state from time *t* to time *t* + *dt*, as indicated in Table 2.

Table 2 shows the Bayes (Berenson, Levine & Krehbiel, 2001) conditional state transition probabilities for the market segment under study, for customers in different states at times *t* and *t*+*dt*. Although interesting in itself, this table does not reflect possible sales of a certain product. Indeed, a customer can stay in the same state but still have made a sales transaction, for instance by replacing one unit with another and disposing of the first unit.

			Time = *t* + *dt*						
			Has zero products		Has one product		Has two products		
			Has never adopted	Had product before	Has one from primary market	Has one from secondary market	Has two products from primary market	Has one product from primary market, one from secondary market	Has two products from secondary market
Time = *t*	Has zero products	Has never adopted	$P(s0 \mid s0(t))$	0	$P(s2 \mid s0(t))$	$P(s3 \mid s0(t))$	$P(s4 \mid s0(t))$	$P(s5 \mid s0(t))$	$P(s6 \mid s0(t))$
		Had product before	0	$P(s1 \mid s1(t))$	$P(s2 \mid s1(t))$	$P(s3 \mid s1(t))$	$P(s4 \mid s0(t))$	$P(s5 \mid s0(t))$	$P(s6 \mid s0(t))$
	Has one product	Has one from primary market	$P(s0 \mid s2(t))$	$P(s1 \mid s2(t))$	$P(s2 \mid s2(t))$	$P(s3 \mid s2(t))$	$P(s4 \mid s2(t))$	$P(s5 \mid s2(t))$	$P(s6 \mid s2(t))$
		Has one from secondary market	$P(s0 \mid s3(t))$	$P(s1 \mid s3(t))$	$P(s2 \mid s3(t))$	$P(s3 \mid s3(t))$	$P(s4 \mid s3(t))$	$P(s5 \mid s3(t))$	$P(s6 \mid s3(t))$
	Has two products	Has two products from primary market	$P(s0 \mid s4(t))$	$P(s1 \mid s4(t))$	$P(s2 \mid s4(t))$	$P(s3 \mid s4(t))$	$P(s4 \mid s4(t))$	$P(s5 \mid s4(t))$	$P(s6 \mid s4(t))$
		Has one product from primary market, one from secondary market	$P(s0 \mid s5(t))$	$P(s1 \mid s5(t))$	$P(s2 \mid s5(t))$	$P(s3 \mid s5(t))$	$P(s4 \mid s5(t))$	$P(s5 \mid s5(t))$	$P(s6 \mid s5(t))$
		Has two products from secondary market	$P(s0 \mid s6(t))$	$P(s1 \mid s6(t))$	$P(s2 \mid s6(t))$	$P(s3 \mid s6(t))$	$P(s4 \mid s6(t))$	$P(s5 \mid s6(t))$	$P(s6 \mid s6(t))$

Table 2. Conditional State Transition Probabilities, from Time *t* to Time *t* + *dt*

Hence, for such models, it is desirable to define a set of actions, such as in Table 3.

a0	Will not buy
a1	Will dispose of one unit (and not buy any)
a2	Will buy exactly one unit on the primary market
a3	Will buy exactly one unit on the secondary market
a4	Will buy exactly two units on the primary market
a5	Will buy exactly one unit on the primary market and exactly one on the secondary market
a6	Will buy exactly two units on the secondary market

Table 3. Definition of Possible Customer Actions

This also makes it possible to define the conditional probabilities for actions *a*0 through *a*6 happening in the time interval]*t.t+dt*], given the fact that a customer can be in any state *s*0 through *s*6 at time *t*. These probabilities do define the possible product sales in the time interval]*t,t+dt*]. Table 4 gives the possible conditional action probabilities for actions *a*0 through *a*6, given the states *s*0 through *s*6. Clearly it is possible to define more states and more actions than the ones given in Tables 1 and 3.

From a manufacturer's point of view, the interesting numbers are the sales in the primary market for a certain product, since they reflect the number of units that have to be produced and that will generate revenue. The manufacturers may be interested in the number of units a customer has already bought, if sales price is dependent on the number of installations for this particular customer.

Aggregate Multi-State Transition or Flow Models

Each of the probabilities *P*(*ai*|*sj*) in Table 4 is a function of time and of the market segment under study. When a set of mutually exclusive market segments are defined, the probabilities for each of these segments can be weighted with the probability of belonging to the segment and summed in order to obtain an aggregate probability *P*(*ai*|*sj*):

$$P(ai(]t,t+dt]\,|\,sj(t)) = \sum_{k=1}^{n} P(ai_k(]t,t+dt]\,|\,sj_k(t)) * P_k \quad (2)$$

where the left hand side reflects the aggregate action probability for action *ai*, given that the customer is in state *sj*, for all possible mutually exclusive market segments *k*. P_k is the probability of belonging to the market segment k, and:

$$P(ai_k(]t,t+dt]\,|\,sj_k(t))$$

indicates the probability for action *ai*, given that the customer is in state *sj* and given that he or she belongs to market segment *k*. The study of these probabilities and the way to segment the market can be referred to as explanatory, since it seeks to explain why and when a member of a certain market segment will commit action *ai*, given that he or she is in state *sj*.

THE BASIC DIFFERENTIAL EQUATION ORIGINALLY PROPOSED BY BASS

Bass (1969) proposed a differential equation to describe the diffusion of a new durable product into the market:

$$\frac{dN}{dt} = \left[p + q\frac{N(t)}{m}\right] * (m - N(t)) \quad (3)$$

where *m* is the market size, *p* is the coefficient of innovation, *q* is the coefficient of imitation and *N*(*t*) is the cumulative number of units placed into the market at time *t*. The left-hand side represents the change in the cumulative number of units per unit of time. There is some logic behind Equation 3 in the sense that the number of units that will be sold is proportional to the remaining market size (m-$N(t)$). The multiplication factor consists of the sum of a constant p and a factor that is proportional to the percentage of the market already "covered" by the new product, for example $N(t)/m$. This proportional factor q therefore represents a factor of imitation, or the influence of the penetrated market upon the unpenetrated market. The constant p therefore represents the percentage of the remaining market that buys the product without being influenced by the installations already in the market place. It is referred to as the coefficient of innovation.

Relation of the Bass Equation to Aggregate Multi-State Flow Models

The book gives very little insight into the derivation of Equation 3, the assumptions that are required in order to accept Equation 3, and the theory behind multi-state flow models. The following represents the author's understanding of the multi-state flow models and will help the reader to understand the implications of adopting the Bass model.

Table 4 gives the conditional probabilities for different customer actions aj, given different possible customer states si. Assuming that the probabilities represent aggregate probabilities over different market segments and that the different states si are mutually exclusive, then the probability $P(a2]t,t+dt])$ that a market member buys exactly one new product in the time interval $]t,t+dt]$ is given by:

$$P(a2]t,t+dt]) = \sum_{i=1}^{n} P(a2]t,+dt] \mid si(t)) * P(si(t)) \qquad (4)$$

where n represents the number of different possible states a customer can be in at time t.

			Time =]*t*,*t*+*dt*]						
			Buys none	Disposes	Buys one unit		Buys two units		
			Will not buy	Will dispose of one unit (and not buy any)	Will buy one unit on the primary market	Will buy one unit on the secondary market	Will buy two units on the primary market	Will buy one unit on the primary market, one on the secondary market	Will buy two units on the secondary market
Time = *t*	Has zero products	Has never adopted	$P(a0 \mid s0(t))$	0	$P(a2 \mid s0(t))$	$P(a3 \mid s0(t))$	$P(a4 \mid s0(t))$	$P(a5 \mid s0(t))$	$P(a6 \mid s0(t))$
		Had product before	0	$P(a1 \mid s1(t))$	$P(a2 \mid s1(t))$	$P(a3 \mid s1(t))$	$P(a4 \mid s0(t))$	$P(a5 \mid s0(t))$	$P(a6 \mid s0(t))$
	Has one product	Has one product from primary market	$P(a0 \mid s2(t))$	$P(a1 \mid s2(t))$	$P(a2 \mid s2(t))$	$P(a3 \mid s2(t))$	$P(a4 \mid s2(t))$	$P(a5 \mid s2(t))$	$P(a6 \mid s2(t))$
		Has one product from secondary market	$P(a0 \mid s3(t))$	$P(a1 \mid s3(t))$	$P(a2 \mid s3(t))$	$P(a3 \mid s3(t))$	$P(a4 \mid s3(t))$	$P(a5 \mid s3(t))$	$P(a6 \mid s3(t))$
	Has two products	Has two products from primary market	$P(a0 \mid s4(t))$	$P(a1 \mid s4(t))$	$P(a2 \mid s4(t))$	$P(a3 \mid s4(t))$	$P(a4 \mid s4(t))$	$P(a5 \mid s4(t))$	$P(a6 \mid s4(t))$
		Has one product from primary market, one from secondary market	$P(a0 \mid s5(t))$	$P(a1 \mid s5(t))$	$P(a2 \mid s5(t))$	$P(a3 \mid s5(t))$	$P(a4 \mid s5(t))$	$P(a5 \mid s5(t))$	$P(a6 \mid s5(t))$
		Has two products from secondary market	$P(a0 \mid s6(t))$	$P(a1 \mid s6(t))$	$P(a2 \mid s6(t))$	$P(a3 \mid s6(t))$	$P(a4 \mid s6(t))$	$P(a5 \mid s6(t))$	$P(a6 \mid s6(t))$

Table 4. Conditional Action Probabilities, in Time Interval]*t*,*t*+*dt*]

From a manufacturer's point of view, the interesting probabilities are those that reflect new product purchases, such as $P(a2)$, $P(a4)$ and $P(a5)$, defined in Table 4. If one assumes that the only two states a customer can be in are $s0$ (has never adopted) and $s2$ (bought one from the primary market), then:

$$\begin{aligned} &P(s0(t)) = 1 - P(s2(t)) \\ &P(a2]t, t+dt] = P(a2]t+dt]\,|\,s0(t))^* \\ &\qquad P(s0(t)) + P(a2]t+dt]\,|\,s2(t))^* P(s2(t)) \end{aligned} \quad (5)$$

If $s0$ and $s2$ are the only possible states a customer can be in at any point in time, then the only actions the customer can perform are either buy ($a2$) or not buy ($a0$). All other actions would place the customer in a different state. Furthermore, a customer cannot buy any new product if he or she already has one; as a consequence, the following relations also hold:

$$\begin{aligned} &P(a2]t, t+dt]) = \frac{dP(s2(t))}{dt} \\ &P(a2]t, t+dt]\,|\,s2(t)) = 0 \\ &P(a0]t, t+dt]\,|\,s2(t) = 1 \end{aligned} \quad (6)$$

Table 5 demonstrates this simplification. The Bass equation assumes that many probabilities equate to zero and that one probability equates to 1.

Using Equation 6 leads to:

$$P(a2]t, t+dt]) = \frac{dP(s2(t))}{dt} = P(a2]t, +dt]\,|\,s0)^*(1 - P(s2(t)) \quad (7)$$

This equation is also referred to as the *hazard rates* and forms the unification basis for the Bass diffusion models and disaggregate-level diffusion models (Roberts, 2000).

At any point in time $P(s2(t))$ is given by:

$$P(s2(t)) = \frac{N(t)}{m(t)}, \quad (8)$$

hence:

$$\frac{d}{dt}\left(\frac{N(t)}{m(t)}\right) = P(a2]t, t+dt]\,|\,s0)^*\left(1 - \frac{N(t)}{m(t)}\right), \text{ or}$$

$$\frac{dN}{dt} = \frac{1}{m}\frac{dm}{dt}N + P(a2]t, t+d]\,|\,s0)^*(m(t) - N(t)) \quad (9)$$

Bass equated $P(a2]t,t+dt]|s0(t))$ to

$$P(a2]t, t+dt]\,|\,s0(t)) = p + q\frac{N(t)}{m(t)}, \quad (10)$$

which leads to:

$$\frac{d}{dt}\left(\frac{N(t)}{m(t)}\right) = \left(p + q\frac{N(t)}{m(t)}\right)^*\left(1 - \frac{N(t)}{m(t)}\right), \quad (11)$$

If $m(t)$ is constant, then this equation reduces to the Bass Equation 3. However, this equation can be considered a generalized Bass equation, which also holds for variable market sizes $m(t)$. In other words, the relevant variable is the market penetration $n(t)=N(t)/m(t)$, irrespective of the market size $m(t)$, and the following equation holds for $n(t)$:

$$\frac{dn}{dt} = (p(t) + q(t)n(t))^*(1 - n(t)). \quad (12)$$

The total cumulative sales $N(t)$, at any point in time is given by:

$$N(t) = m(t)^* n(t). \quad (13)$$

The generalized Bass Equation 11 can be rewritten as:

$$\frac{dN}{dt} = p(t)m(t) + (\frac{1}{m}\frac{dm}{dt} + q(t) - p(t))N(t) - \frac{q(t)}{m(t)}N(t)^2 \qquad (14)$$

Also p and q are considered constant over time, then there is an analytic solution for Equation 14, as described in Appendix A.

Time-Dependent Parameters—Riccati Differential Equation

Equation 14 is of a general form also known as the Riccati differential equation (Boyce & DiPrima, 1977). In Appendix A it is shown that $N(t)$ is given by:

$$N(t) = \frac{m(t)}{q(t)}\frac{1}{Y(y)}\frac{dY}{dt}, \qquad (15)$$

with $Y(t)$ a solution of the homogeneous linear differential equation with non-constant coefficients $p(t)$ and $q(t)$:

$$\frac{d^2Y}{dt^2} + \left[(p(t) - q(t)) - \frac{dq}{dt}\right]\frac{dY}{dt} - pqY = 0 \qquad (16)$$

Note that $Y(t)$ is independent of $m(t)$, which implies that the diffusion process is independent of $m(t)$ and that $N(t)$ is simply proportional to $m(t)$, in consistency with Equations 13 and 15.

		Time =]$t,t+dt$]						
		Buys none	Disposes	Buys one unit		Buys two units		
		Will not buy	Will dispose of one unit (and not buy any)	Will buy one unit on the primary market	Will buy one unit on the secondary market	Will buy two units on the primary market	Will buy one unit on the primary market, one on the secondary market	Will buy two units on the secondary market
Time = t: Has zero products	Has never adopted	$P(a0\|s0(t))$	0	$P(a2\|s0(t))$	0	0	0	0
	Had product before	0	0	0	0	0	0	0
Has one product	Has one product from primary market	1	0	0	0	0	0	0
	Has one product from secondary market	0	0	0	0	0	0	0
Has two products	Has two products from primary market	0	0	0	0	0	0	0
	Has one product from primary market, one from secondary market	0	0	0	0	0	0	0
	Has two products from secondary market	0	0	0	0	0	0	0

Note: Assuming that $s0$ and $s2$ are the only possible states a customer can be in implies that $P(a0]t,t+dt]|s2(t))=1$ and $P(a2]t+dt]|s2(t))=0$

Table 5. Conditional Action Probabilities, in Time Interval]$t,t+dt$], for the Bass Equation

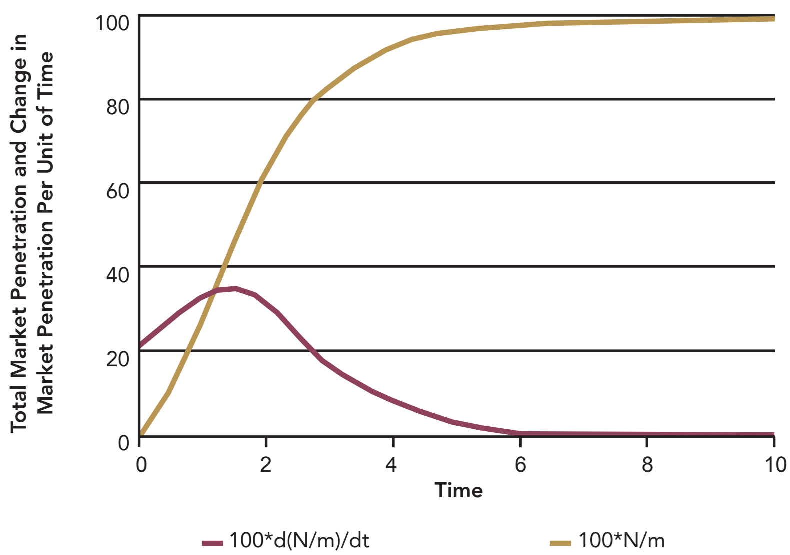

Figure 1. Cumulative and Differential Diffusion Population, p=0.2, q=0.9

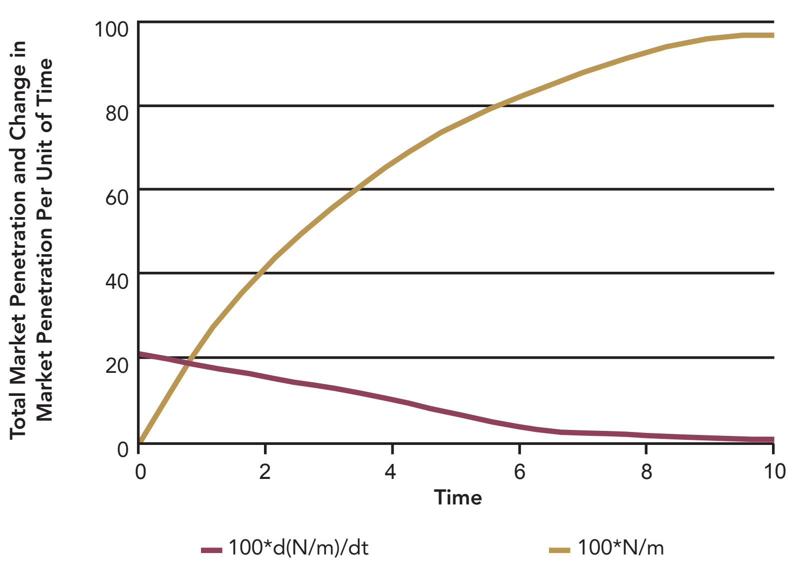

Figure 2. Cumulative and Differential Diffusion Population, p=0.2, q=0.2

If p and q can be considered constant, the solution for cumulative sales $N(t)$ is given by the following (see Appendix A):

$$N(t) = m(t)\frac{1-e^{-(p+q)t}}{1+\frac{q}{p}e^{-(p+q)t}} \quad . \tag{17}$$

Evaluating Equation 17 at two different times and subtracting the resulting two numbers from one another yields the total sales in this time period. Evaluating Equation 17 at the beginning and the end of the year and subtracting the two numbers from one another yields the annual sales data.

In other words, an annual sales forecast is given by:

$$s(t) = N(t+1) - N(t)$$

which also can be used as an approximation of dN/dt, as given by Equation 14.

Examples of Solutions for Different Parameter Combinations

Figures 1 through 5 show some examples of the total market penetration or the relative cumulative diffusion population $N(t)/m(t)$ and change rate per unit of time of the market penetration

$$\frac{d}{dt}\left(\frac{N(t)}{m(t)}\right)$$

, for different constant parameters p and q.

$$\frac{d}{dt}\left(\frac{N(t)}{m(t)}\right)$$

also can be designated as the rate of change of the diffusion population, relative to the total market size $m(t)$.

Five different combinations of p and q were chosen:

p=0.20, q=0.90
p=0.20, q=0.20
p=0.90, q=0.20
p=0.90, q=0.90
p=0.01, q=0.60

It can be seen from Figures 1 through 5 that the evolution over time of the cumulative penetration $N(t)/m(t)$ and the penetration change per time unit $d(N/m)/dt$, are affected considerably by the parameters p and q.

Given the annual sales data of a certain product, the question arises whether one can determine the parameters p, q, and m and then use these parameters to make a sales forecast. Such forecasts are very useful in performance capacity planning of a production plant, estimating unit production costs, forecasting revenues and cash flows over time, and calculating the stock price of companies given their product range, predicted sales numbers, and unit cost.

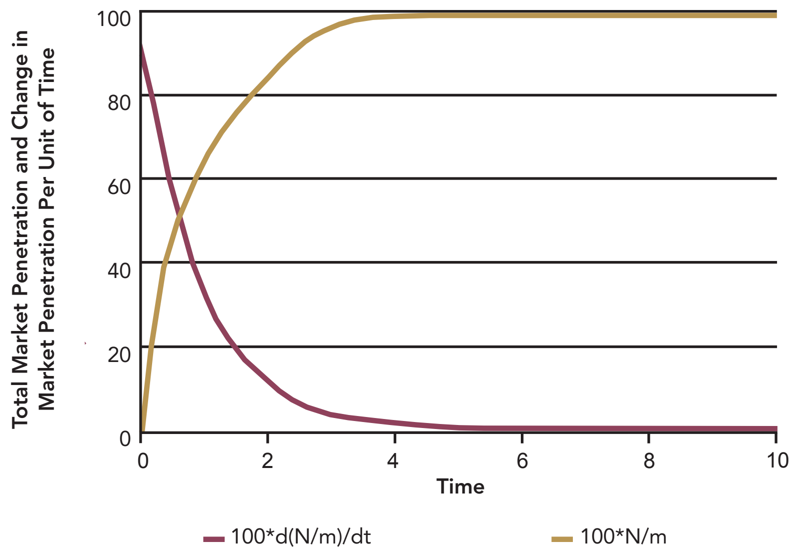

Figure 3. Cumulative and Differential Diffusion Population, p=0.9, q=0.2

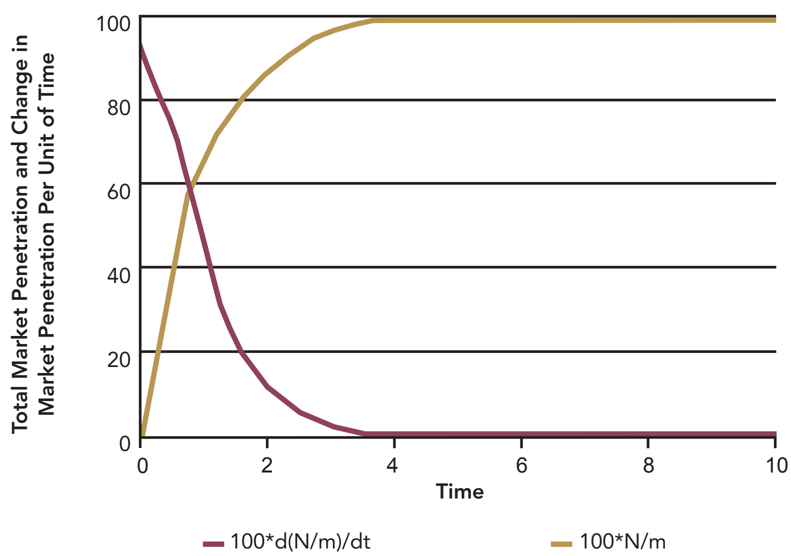

Figure 4. Cumulative and Differential Diffusion Population, p=0.9, q=0.9

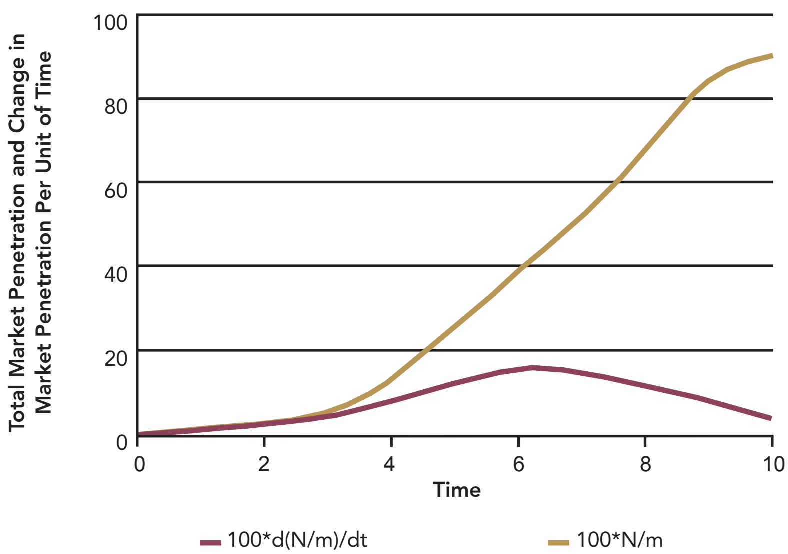

Figure 5. Cumulative and Differential Diffusion Population, p=0.01, q=0.6

Parameter Extraction Examples

From a managerial point of view, the extraction of the diffusion parameters early in the marketing process is the most interesting for planning purposes. From an academic point of view, the "after the facts" determination of the diffusion parameters is interesting for evaluating the validity of the diffusion model and possibly refining the model. From the managerial point of view, the latter is interesting for as a "lessons learned" exercise in order to incorporate improvements into subsequent forecasts.

Early in the Diffusion Process

The earlier it is in the diffusion process, the more difficult parameter extraction is. One possible solution is to use the parameters for a comparable product in the same market. The older product preferably should be comparable in cost and benefits for the user, yet should be different enough from the new product that the new product is not purchased to replace the older one.

Another possibility is to use techniques similar to the ones described in the next paragraph. This usually leads to larger errors in the parameters and therefore in the forecasts.

After the fact, the parameters, *p*, *q* and *m* can be found by minimizing an error function between the predicted cumulative sales, as given by Equation 17 sub-sampled annually, and the actual annual sales. The choice of the error function will have an influence on the values of *p*, *q* and *m*. It is therefore important to choose an appropriate error function.

An obvious error function is the sum of the squared errors of the cumulative annual sales:

$$E_k^2 = \sum_{i=1}^{k} (N_i - S_i)^2, k = number\_of\_available\_years\_of\_sales\_data, \quad (18)$$

in which N_i is the cumulative predicted sales data for year *i*, as given by Equation 17, S_i is the actual cumulative sales data, and *k* is the number of available years of sales data. For three years of available sales, the error function E_k, can be made zero, with at least one combination of *p*, *q*, and *m*, since this would lead to a set of three non-linear equations and three unknowns. Minimization of E_k, with respect to *p*, *q*, and *m* requires that *p*, *q*, and *m* are solutions of the following set of equations:

$$\sum_{i=1}^{k} (N_i - S_i)\frac{\partial N_i}{\partial p} = 0$$
$$\sum_{i=1}^{k} (N_i - S_i)\frac{\partial N_i}{\partial q} = 0$$
$$\sum_{i=1}^{k} (N_i - S_i)\frac{\partial N_i}{\partial m} = 0 \quad (19)$$

Given the nature of N_i, this is a set of non-linear equations in *p*, *q* and *m*, which can have several solutions, also referred to as local minima of E_k. It is also clear that the lower the number of years of sales data available, the more local minima will exist. Note that a company may be able to use cumulative monthly sales data to determine a first estimate of *p*, *q*, and *m*. One would expect that a minimization based on cumulative monthly sales data might make possible a more accurate determination of p, q, and m in a shorter time period. However, there is some danger in this way of working, because monthly data may exhibit seasonal effects or other larger relative variations.

Other error functions may be used, such as:

$$D_k^2 = \sum_{i=1}^{k} (N_{i+1} - N_i - s_i)^2, k = number\_of\_available\_years\_of\_sales\_data \quad (20)$$

In Equation 20 N_{i+1}-N_i is the forecast of the annual sales in year i, whereas s_i is the actual annual sales of year i.

It is expected that the error functions cients p, q and m.

Yet another alternative error functions E_k and D_k would yield rather similar results for the coefficients p, q, and m.

Yet another alternative error function

$$F_k = (N_k - S_k)^2,$$

which requires the cumulative sales data to match at only one point in time, is expected to lead to a less accurate estimate of p, q, and m.

In any case there are a few conditions that need to be imposed on the minimization process, such as

$$\begin{aligned} N_k &\geq S_k \\ p &> 0 \\ q &> 0 \end{aligned} \quad (21)$$

Appendices B and C discuss some alternative ways of looking at the problem.

Examples from the Printing Industry

XEROX DOCUTECH FAMILY OF PRODUCTS

Limitations

The discussion that follows is made based on an assumption about market size, which was from 11 years of available sales data. Of course, this assumption might be invalid, which would make some of the conclusions obsolete. The estimated market size is dependent on the sales evolution during this 11-year period. If, for instance, the sales data had been affected by operational decisions within the company, then the derived market size would be invalid. This has significant implications for the discussion and basically invalidates some possible conclusions. As an example, Xerox decided to reorganize its sales force in 1998, which can have a significant impact on sales.

As a fundamental limitation it will be shown that the Bass diffusion model presumes an accurate knowledge of the market size. Other limitations of the Bass diffusion model mentioned in earlier paragraphs, apply to this example as well.

Discussion

The Xerox DocuTech family of products (DT135, DT6135, DT6180, DT6100, DT6115, DT6155) is considered for this example. This group of products was truly innovative in 1990, allowing sophisticated, high-speed reproduction of documents, both in hard copy and in electronic form. The products offered quality, ease of reproduction, and user friendliness through network connectivity and in-line finishing. The underlying technology can be considered disruptive (Ettlie, n.d.), in that it initiated some interesting market dynamics that have affected the field of printing since its introduction and will continue to do so for years to come. This type of product in the mean time has entered the field of digital color printing and will continue to change the market in this area as well.

Annual and cumulative sales data (Holt, n.d.) are given in Table 6 for all of the products in this family, in number of units.

Year	*Annual* Sales s_i in units	*Cumulative* Sales s_i in units
1990	99	99
1991	1,047	1,146
1992	1,809	2,955
1993	1,783	4,738
1994	2,293	7,031
1995	2,441	9,472
1996	2,919	12,391
1997	3,310	15,701
1998	3,878	19,579
1999	3,653	23,232
2000	3,124	26,356

Note: The total sales for the last two years were communicated to the author in extremis: 2401 (2001) and 2190 (2002). Predicted sales for this time period, based on six years of sales data and a market size of 38,833, are 2,591 (2001), 2,051 (2002).

Table 6. Sales Data for Xerox DocuTech, 1990-2000

A simple Bass diffusion model in a continuous form (see Appendix B for a discussion on a continuous versus a discrete model:

$$\frac{dN}{dt} = \left[p + \frac{q}{m} N(t)\right] * (m - N(t)) \qquad (22)$$

has been applied to this data set, with estimates for p, q and m based on 3, 4, 6, 8, and 11 years into the marketing process. The solution to the Bass diffusion equation is given by the following (see Appendix A):

$$N(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \qquad (23)$$

The parameters *m*, *p* and *q* are estimated five times, using 3, 4, 6, 8 and 11 years of sales data, by minimizing the sum of the square errors of actual cumulative sales versus forecasted cumulative sales:

$$E_k^2 = \sum_{i=1990} (N_i - S_i)^2, k = 1992, 1993, 1995, 1997, 2000 \qquad (24)$$

Where N_i is the forecast for the cumulative sales as given by Equation 23 and S_i is the actual cumulative sales of year *i*, listed in Table 6.

The constraint on the cumulative market size *m*, is that it has to be larger than the cumulative sales up to that point in time. It was found that Equation 24 has more different local minima for smaller values of *k*, and that for a few market sizes *m*, a pair of parameters *p* and *q* can be found that "traps" the minimization algorithm into a local minimum.[1] Appendix C gives some examples of the solutions that were obtained for different numbers of years of available sales data and different initial estimates for the parameters, *p*, *q*, and *m*. It can be seen in Appendix C that the final market size found by the Solver follows the initial guess of the market size rather closely for three and six years of available sales data. For 11 years of available sales data, the market size converges, in most cases, to the same number.

The following analysis was made by using the market size obtained for 11 years of available sales data as an initial guess for the market size for four cases with fewer years (3, 4, 6, and 8 years) of available sales data.

Figure 6 shows the result of the five different forecasts of the cumulative sales, based on the five estimates for *p*, *q* and *m*. The dotted blue line indicates the actual cumulative sales from 1990 to 2000, as given in Table 6. The colored lines indicate the forecasts based on 3, 4, 6, 8, and 11 years of sales data, for *p*, *q*, *m* parameters that minimize E_k (k=3, 4, 6, 8, 11), given by Equation 24. Similarly, Figure 7 shows the corresponding annual sales data in number of units per year.

The values of *p*, *q*, and *m* that minimize E_{11}, are *p*=0.015, *q*=0.343, and *m*=38,833. The validity of these parameters will be discussed later in this paper.

Forecasted annual sales, depicted in Figure 7, are given by Equation 22 for the parameters *p*, *q*, *m* that minimize E_k. Actual sales are shown as the light gold line and are given in Table 6.

It is interesting to note that the forecasts based on three and four years of sales data are not very accurate in terms of the annual sales volume prediction. A forecast based on five years of sales also was not very accurate. (This forecast is not shown in Figures 6 or 7, for the sake of clarity, but lies between the forecasts based on four and six years of data.)

The forecast based on six years (1990 to 1995) of sales data is surprisingly accurate, over a five-year time horizon into the future, from 1996 until 2000. If additional years of sales data are used to determine *p*, *q*, and *m*, the resulting forecasts do not change substantially; they merely oscillate closely around each other and around the actual sales data.

Forecasts differ mainly in the annual sales distribution, as can be seen from Figures 6 and 7. Forecasts based on few years of available sales data predict too-high sales early in the marketing process and too-low sales later in the marketing process. Sales forecasts based on more years of available data correct for the early overestimation and late underestimation errors. Note that such estimation errors would affect

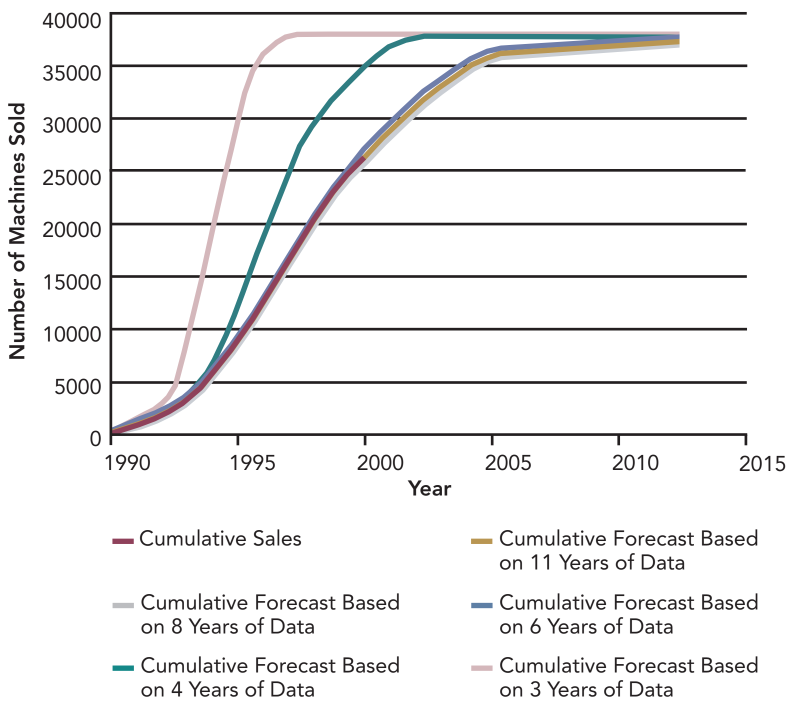

Figure 6. Cumulative Sales of DocuTech Family and Forecasts With Minimal Cumulative Error, Based on 'n' Years of Sales Data

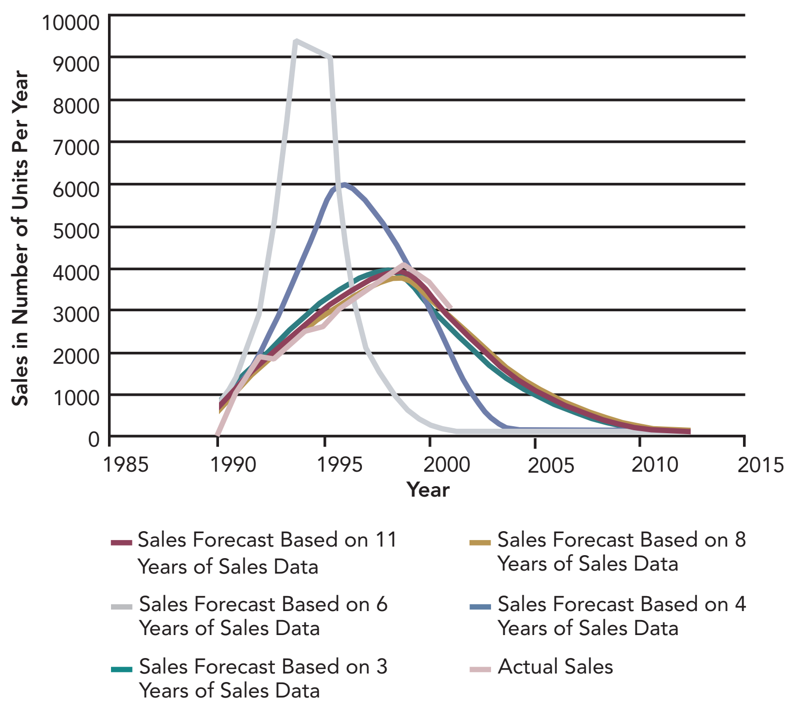

Figure 7. Predicted Annual Sales of DocuTech Family in Number of Units, Based on Different Number of Years of Available Sales Data

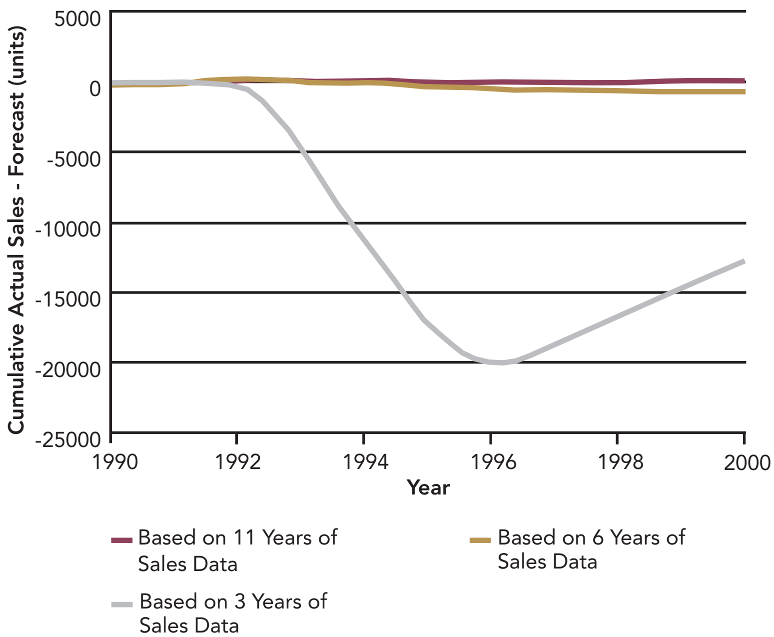

Figure 10. Cumulative Sales Forecast Error

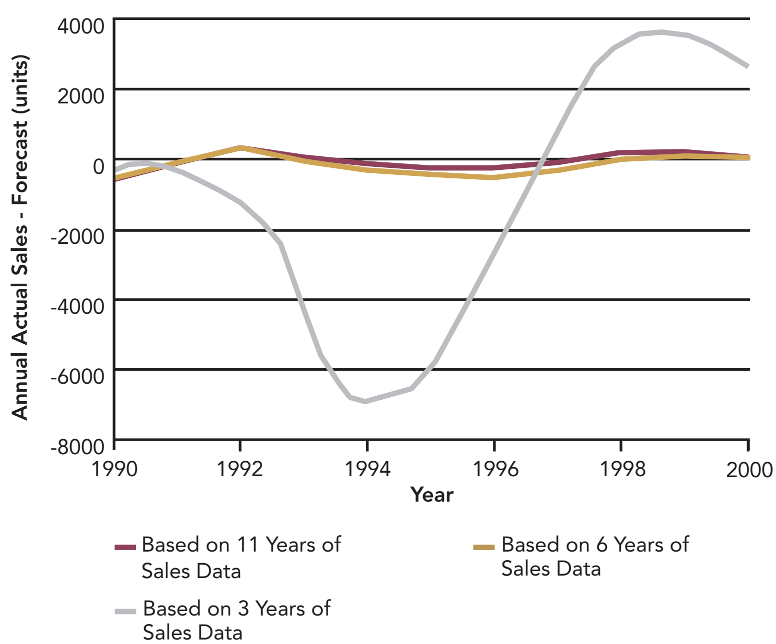

Figure 11. Annual Sales Forecast Error

the cash flow distribution and thus the net present value of the cash flow and the stock prices.

Figures 8 and 9 show the values of the Bass innovation coefficient p, and imitation coefficient q, for forecasts based on 3 to 11 years of available sales data and an initial estimated market size of around 39,000 units.

As can be seen, forecasts based on fewer years of sales data overestimate the coefficient of imitation q by as much as 66% and underestimate the coefficient of innovation p by as much as 200%. It can also be observed from Figures 8 and 9 that the p and q parameters are almost constant for the last six forecasts, based on 6, 7, 8, 9, 10, and 11 years of available sales data. This increases the confidence in the model, which assumes that m, p, and q are time-independent parameters.

The market size forecast is very similar for all estimates, based on any number of years of available sales data, with a maximum difference of about 6.5%. This is a consequence of the fact that, for any of the p, q, and m estimation attempts, the iteration process was started with an initial market size m, as found when using 11 years of actual sales data. In Appendix C it is shown that the error function E_k has many local minima and that the value of p and q that minimize E_k are dependent on the estimated market size m.

From a managerial viewpoint it is important to obtain an accurate idea about market size early in the marketing process, such that capacity planning and supply chain management can be done efficiently and reasonable profitability forecasts can be made. To emphasize this point, the information presented in Figures 6 and 7 is represented as forecasting error graphs in Figures 10 and 11.

Forecasts based on three years of sales data lead to significant differences between the actual sales and the forecasted sales, cumulative as well as annual, emphasizing the difficulty of obtaining accurate sales forecasts and total market size with only a few years of sales data available.

It is noteworthy that the cumulative number of installations of the Heidelberg Digimaster is around 3,500.[2] This product is probably the closest competitor to the Xerox DocuTech family. This number does not substantially change the estimated market size.

Comparison of the Parameters *p* and *q* With Other Products

The numbers for p and q obtained in the previous paragraphs for the DocuTech family of products can be compared with some of the available numbers found in the literature for production technology equipment. Table 7 was constructed using data from new product diffusion models (Lilien et al., 2000).As can be seen, the p and q parameters are comparable, yet different enough for these different types of products.

Example Conclusion

From this discussion, it is concluded that the Bass diffusion model makes it possible to estimate the market size m if a sufficient number of years of sales data are available. If only a few years of sales data are available it is much more difficult to predict the market size. This shows that there is a need to complement the Bass diffusion model with a model for determining the market size. Indeed, the market size has a significant influence on the estimation of the diffusion parameters p and q.

The market size used in this example was estimated using 11 years of sales data by minimizing the cumulative error function in Equation 24 for the parameters p, q, and m. An indication that this market size could be correct is provided by the fact that forecasts based on six years of available sales data, using this market size, do predict sales accurately (with a 5% to 10% error) for seven years into the future, from 1996 until 2002. This is quite remarkable.

Furthermore, it was shown that even when the market size is known the parameters p and q of the diffusion model, obtained through minimization of the error function in Equation 24, are quite dependent on the number of years of available sales data. This has implications for the usability of the Bass diffusion model early in the marketing process.

Assuming that the market size determined by using 11 years of sales data is correct, it could be concluded that the sales for black-and-white DocuTech like products will continually decline in the future, and will drop to zero at around 2010. In this time frame the remaining market size is estimated to be around 8,500 units. This conclusion also assumes that there are no replacement purchases and no multiple product purchases. It is anticipated that there will still be a market for replacement units after 2010 and that the total opportunity is therefore larger than 8,500 units. It is also anticipated that there is a global opportunity that has not yet been exploited because of too-recent political and macro-economic changes in parts of the world.

Product Type	Period of Analysis	*p*	*q*	Penetration Data *n* (%)
Oxygen Steel Furnace (USA)	1955-1980	0.002	0.435	60.5
Oxygen Steel Furnace (France)	1961-1980	0.008	0.279	88.4
Oxygen Steel Furnace (Japan)	1959-1975	0.049	0.333	81.3
Steam Versus Merchant Ships (UK)	1815-1965	0.006	0.259	86.7
Plastic Milk Containers (1 gallon)	1964-1987	0.02	0.255	100
Plastic Milk Containers (0.5 gallon)	1964-1987	0.00	0.234	28.8
Xerox DocuTech Family (estimated primary market size 38,833 units)	1990-2002	0.015	0.346	80.0
Digital Color Presses (see example in Color Copiers and Digital Color Presses section, this data assumes a primary market size of 120,000 units)	1993-2002	0.0039	0.433	40.0

Table 7. *p* and *q* Parameters for Different Production Technology Equipment

Replacement purchases and additional sales of new products with enhanced capabilities to existing customers are not included in the basic Bass model. Models have been developed to deal with replacement and multiple purchases. An overview of such models can be found in new product diffusion models (Ratchford et al., 2000). The model described by B. T. Ratchford uses discrete time intervals rather than a continuous time model such as the Bass model. Replacement purchases in this model are given by:

$$r_t = \sum_{k=1}^{t-1} [Q(k-1) - Q(k)] s_{t-k} , \qquad (25)$$

where $Q(k)$ represents the percentage of units that survive for a time longer than k. Thus, the percentage of products that break down in their k-th year is given by $Q(k\text{-}1)\text{-}Q(k)$. Products sold at a time $t\text{-}k$ have a lifetime k at time t. The total number of sales made at time $t\text{-}k$ is denoted as s_{t-k}. Hence, the total number of products that irreparably break down by the end of the period t and were sold at time $t\text{-}k$ is given by $(Q(k\text{-}1)\text{-}Q(k))*s_{t-k}$. Equation 25 assumes that all of these units will be replaced by the end of time period t. Since many of the original DT135 units are still operational in the field, it is assumed that a minority of sales has been replacement purchases.

Of course, customers now have a broader choice of models from competitors with which to replace the old units. Customer satisfaction is therefore a crucial parameter for the replacement purchase process.

Models also have been developed for dealing with multiple-product interactions (Bayus et al., 2000) and competition (Chaterjee et al., 2000).

Clearly, one can refine the basic Bass diffusion model and introduce more complexity.

COLOR COPIERS AND DIGITAL COLOR PRESSES

Limitations

The following example is based on the aggregate sales of different types of digital color presses and color copiers with different levels of performance and different characteristics. It is not clear that these machines are sold in the same market. An analysis based on aggregate sales offers an advantage over the sales of a single vendor, since the numbers are less sensitive to company internals. No official confirmation of the sales numbers could be obtained from the different vendors to increase the credibility of the numbers.

Discussion

A second example is based on data obtained from Frank Romano of RIT and was published in TrendWatch Graphic Arts (2003, May). The table in Appendix D lists extended data and the speeds of each of the presses. The cumulative and summarized version used in this example is given in Table 8. The table reports sales of different models from different vendors (Canon, Heidelberg, HP, Xeikon, Xerox and others). The simple Bass diffusion model was applied to this data set.

Year	Sales (in units)	Cumulative Sales (in units)
1993	31	31
1994	295	326
1995	626	952
1996	1,550	2,502
1997	4,239	6,741
1998	6,075	12,816
1999	6,175	18,991
2000	8,320	27,311
2001	9,220	36,531
2002	10,910	47,441

Table 8. Cumulative and Summarized Digital Color Press Installations, 1993-2003

Figure 12 shows the forecasts of cumulative sales based on 3, 4, 6, 8, and 10 years of available sales data. A similar method was used to estimate the market size at slightly over 71,000 units. This means that 10 years of available sales data were used, in combination with minimizing the error function in Equation 24 for the parameters *p*, *q* and *m*. It was again found that this leads to a relatively robust market size, independent of the initial guess of *p*, *q*, and *m*, but it was less robust than that for the DocuTech example. As a consequence, this market size requires more investigation.

The market size was therefore also estimated using the parameters p=0.015 and q=0.346, which were the ones found for the 11 years of DocuTech data. The market size *m* was estimated by minimizing E_{10} (Equation 24), by varying only the market size *m*. In this case the market size was found to be close to 79,000, about 10% higher than the market size found with the previous method. It should be noted that this market size is double that found for the Xerox black-and-white products.

From Figure 12 it can also be seen that the estimates based on 8 and 10 years of available sales data are very close to one another. The estimates based on fewer years of available sales data are a little less accurate, but they are still more accurate than in the example discussed in the "Examples From the Printing Industry" section of this paper.

Figure 13 shows the actual annual sales and the sales rate dN/dt, which is a good approximation of the annual sales forecast.

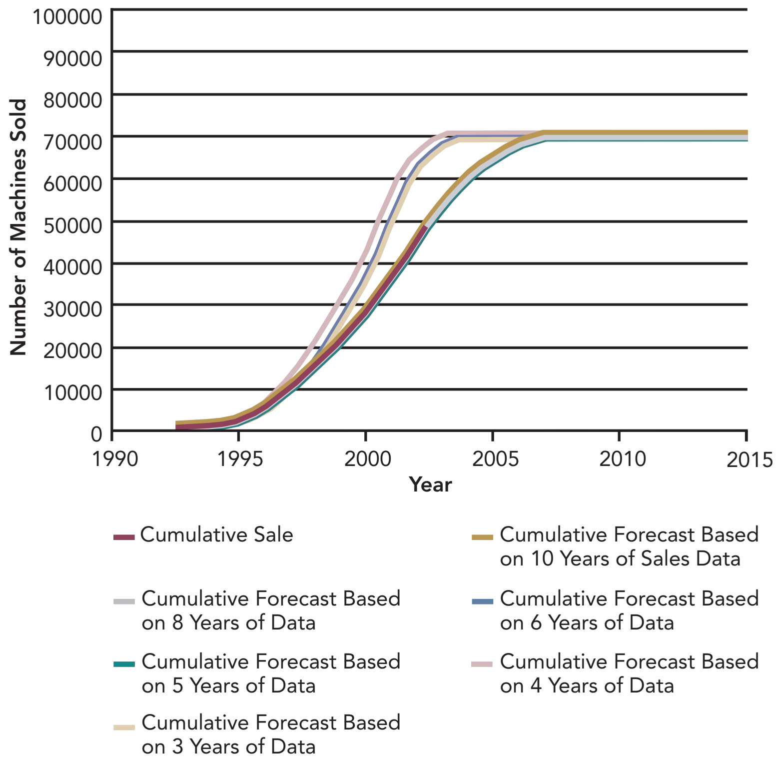

Figure 12. Cumulative Sales of Color Copiers and Presses and Forecasts with Minimal Cumulative Error, Based on 'n' Years of Sales

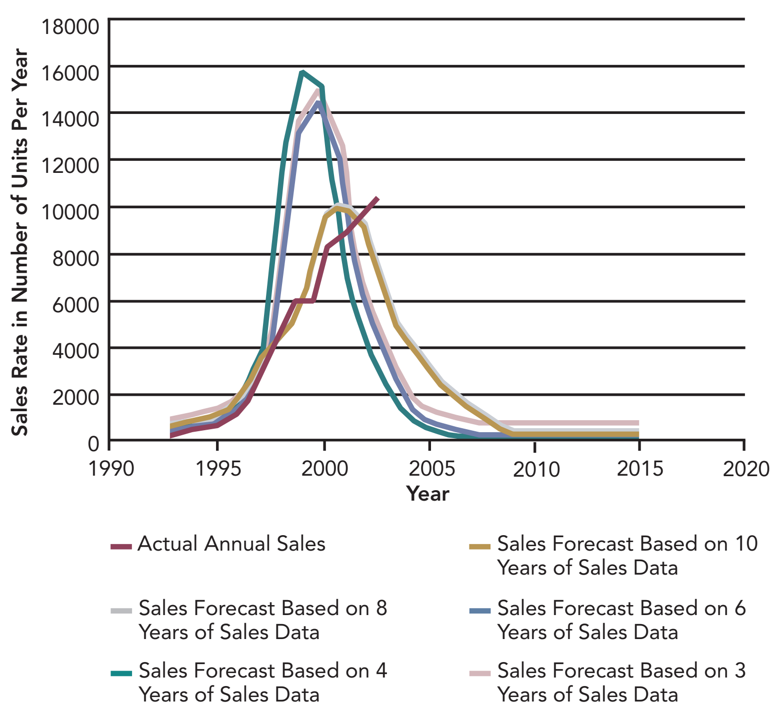

Figure 13. Predicted Sales Rate of Color Copiers and Presses in Number of Units, Based on Different Number of Years of Available Sales Data

As a comparison the alternative error function in Equation 20 was used to determine the *p*, *q* and *m* parameters for the case where 10 years of sales data are available. Table 9 shows the comparison between the *p*, *q* and *m* parameters when error functions E_k (Equation 18) and D_k (Equation 20) are used.

	Error Function E_k	Error Function D_k
p	31	31
q	295	326
m	626	952

Table 9. Comparison of Error Functions

As can be seen in Table 9, the *p*, *q*, and *m* parameters are close to one another, regardless of which error function is minimized. As a consequence, the forecasted sales are also rather similar in both cases. Figure 14 shows the forecasted annual sales data versus the actual sales data, which can be compared to Figure 13, and illustrates again that minimizations of the error functions D_k and E_k yield very similar forecasts.

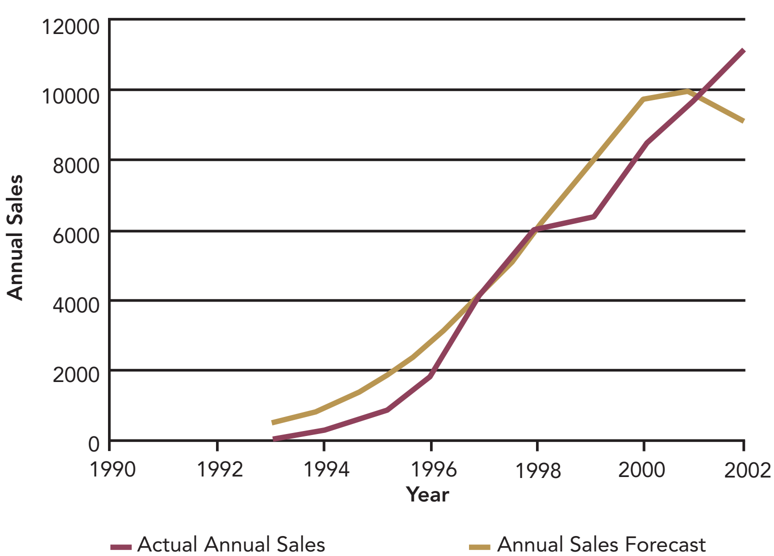

Figure 14. Annual Sales: Actual Versus Forecast, Through Minimization of the Summed Square Error of Annual Sales

Figure 14 shows a declining sales forecast in the last year, while the actual sales data is still showing an increase. This again suggests that the estimated market size found by minimizing the error function E_k should be viewed with caution, as should the validity of the parameters shown in Table 9.

The parameters *p*, *q*, and *m* that were found by minimizing the different error functions E_k for the different numbers of years of available sales data are shown in Figures 15 and 16.

As can be seen in Figures 15 and 16, it is not obvious that the parameters *p*, *q*, and *m* can be considered constant over time for any number of years of available sales data. The variation in the coefficients is smaller in the last three years. Also, these coefficients are a little different from the *p* and *q* numbers found in the Xerox DocuTech example, although not dramatically different, as will be discussed in more depth in the next paragraph.

In order to assess the consequences of the different estimates for the parameters *p*, *q*, and *m*, a comparison is made using the *p* and *q* parameters for the Xerox DocuTech products for the color products. The market size that would be required to have the same diffusion for the color presses as for the DocuTech products was found by fixing p=0.015 and q=0.346 and minimizing E_{10} with respect to *m*. This leads to a market size of 79,000 units.

Figure 17 shows the estimated cumulative sales under these circumstances, compared to the previous estimates and the actual sales.

Clearly, the parameters *p* and *q* found for the DocuTech products cause the forecast for the color presses to deviate more from the actual sales than the previous forecasts, and therefore it can be stated that the diffusion parameters are definitely different for the color products than for the black-and-white products.

As the next step, an initial market size of 79,000 units was chosen. Using this market size and optimizing for *p*, *q*, and *m* in order to minimize E_{10} leads to values for *p* and *q* that are closer to the values shown in Figures 15 and 16. The results of the forecast for this market

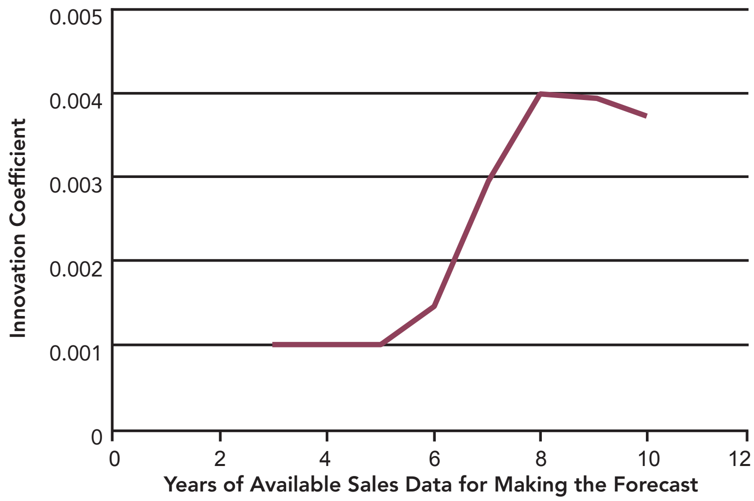

Figure 15. Coefficient of Innovation (p) of Bass Model

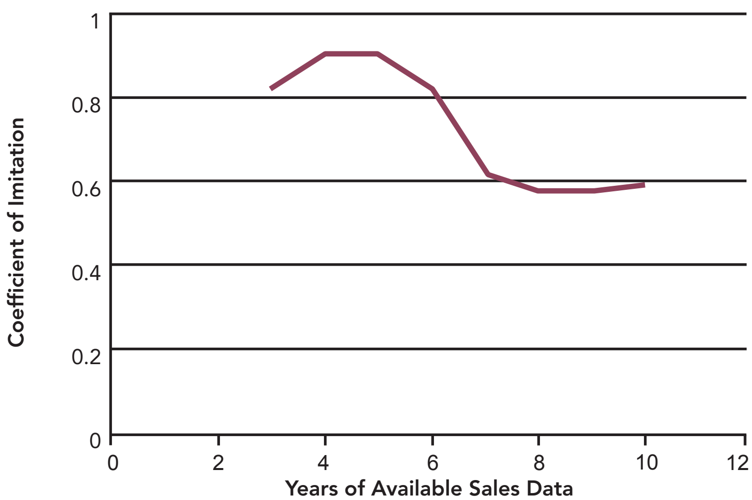

Figure 16. Coefficient of Imitation (q) in the Bass Model

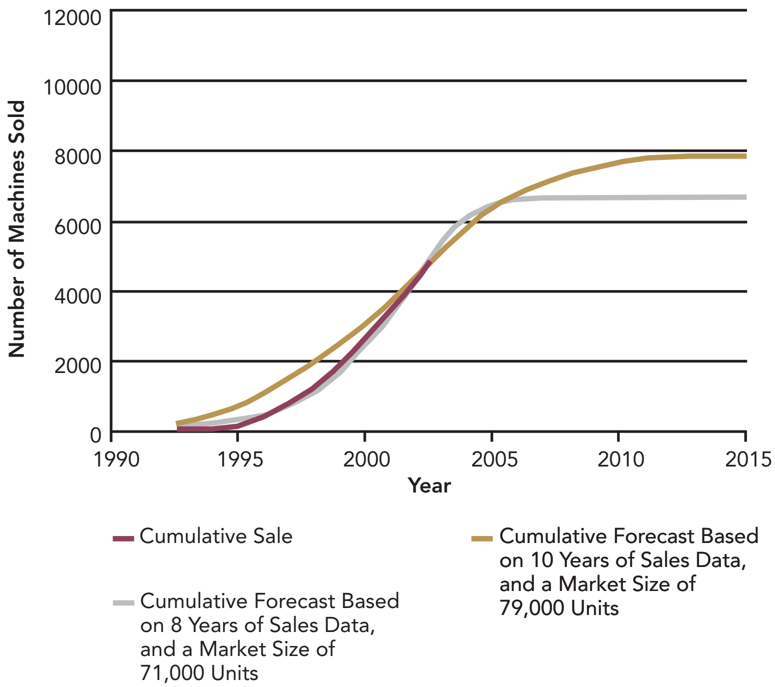

Figure 17. Cumulative Sales of Color Copiers and Presses and Forecasts with Minimal Cumulative Error, Based on 'n' Years of Sales

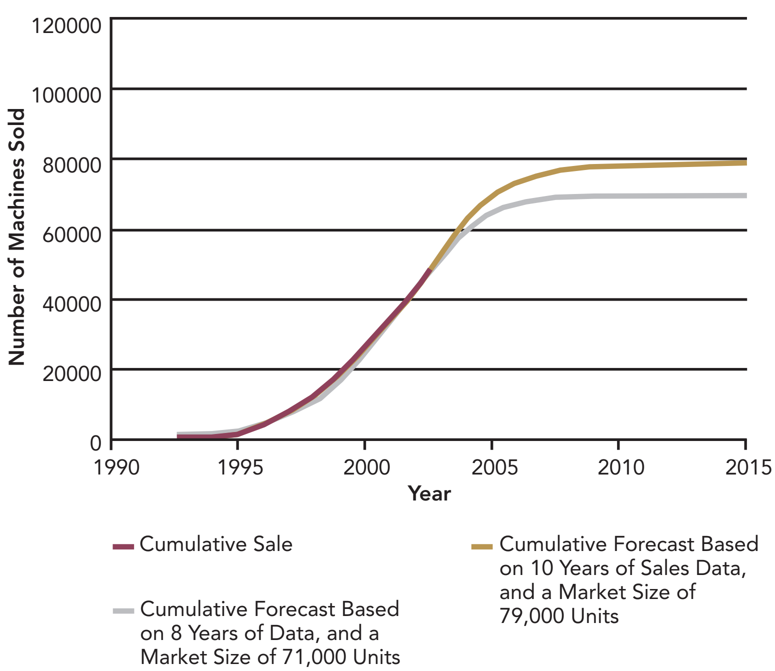

Figure 18. Cumulative Sales of Color Copiers and Presses and Forecasts with Minimal Cumulative Error, Based on 'n' Years of Sales

size are shown in Figure 18, again in comparison to the previous sales forecasts and the actual sales. As can be seen from Figure 18, the forecast more closely matches the actual sales.

Finally, in order to demonstrate the difficulty in estimating the market size using the Bass diffusion model, the market size was assumed to be 120,000 units, Figure 19 shows the obtained forecast against the actual sales and the previous forecasts.

The parameters p and q corresponding to m=120,000 are p=0.0039 and q=0.433. Clearly p is about the same, and q is quite a bit lower than for a market size of 71,000 units, for which p=0.0036 and q=0.5742.

In any case, it can be seen from Figures 17 to 19, that the majority of the sales will have occurred by 2010, for both the conservative market size of 71,000 units and the more aggressive market size of 120,000 units. This implies that an additional 20,000 to 70,000 units would be sold in this time frame in the primary market, excluding replacement units.

The question arises: what is the true market size? A possible further investigation might use the approach given in Appendix C to determine a better approximation for the market size of digital color presses. The U.S. Census Bureau data (n.d.) shows that there are about 42,000 establishments in the NAICS code 323, which refers to the printing industry and its related services. The European community most likely has a similar number of establishments. This excludes important markets such as Japan, Asia, Australia, and the part of Europe not belonging to the EC. A market size of 120,000 units could therefore be a feasible number.

Example Conclusion

Conclusions similar to the ones made for the previous example can be made for this example.

- The estimation of market size works relatively well when more years of sales data are available, yet the market size could be much larger than 70,000 units. The majority of the sales will happen between now and 2010, for market sizes up to 120,000 units

- The *p* and *q* parameters seem to be different from the ones found for the Xerox DocuTech family of products, and using the DocuTech parameters for the color market causes quite large errors between the forecast and the actual sales. The color product diffusion is faster than the black-and-white product diffusion, mainly because of a higher factor of imitation *q*.

- The market size has a considerable influence on the parameters *p* and *q*.

- The parameters *p* and *q* depend on the number of years of available sales data, even if the same market size is used in all cases.

- In the case of the color market the forecasts based on 6 or fewer years of sales data were found to deviate somewhat from the forecasts based on 8 and 10 years of available sales data.

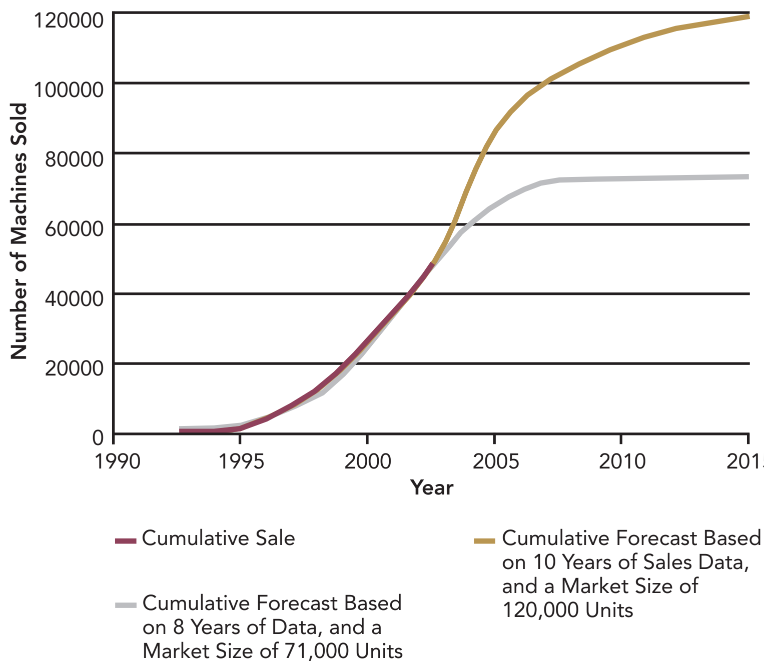

Figure 19. Cumulative Sales of Color Copiers and Presses and Forecasts with Minimal Cumulative Error, Based on 'n' Years of Sales

Conclusion

This paper used the basic Bass diffusion model to investigate the diffusion of digital black-and-white and digital color presses in the printing industry. It was found that digital color presses diffuse faster in the marketplace than digital black-and-white presses, mainly because of a stronger imitation factor.

From this investigation, it is concluded that the primary markets for these digital presses, excluding replacement purchases, will be penetrated almost completely by 2010 to 2015, assuming that the market size is smaller than 120,000 units. It is anticipated that an additional 30,000 to 80,000 units, black-and-white as well as color, will be installed at that point. The market potential will be even larger than that when the replacement purchases are included.

It is also concluded that the basic Bass diffusion is remarkably capable of predicting sales for many years into the future, once 6 to 10 years of sales data are available. For the black-and-white digital presses, the forecasting accuracy was found to be between 5% and 10% for seven years into the future, using only six years of sales data and a good estimate of the market size.

Future work on this topic might include similar analysis and model development with alternative minimization criteria to better estimate market size. One such possibility would be to find the market size m, which minimizes the variation in p and q over time, subject to the condition that p and q are determined by minimizing the error function (Equation 24). Other possibilities for research would be to use time-dependent diffusion parameters. The main purpose of such models would be twofold, first to establish techniques for reliably estimating the market size from the available data, and second to reduce the number of years required to make an accurate forecast of the sales evolution.

References

Bass, F. M. (1969). A new-product growth model for consumer durables. *Management Science, 15* (January), 215-227.

Bayus, B. L., et al. (2000). Growth models for multi-product interactions: Current status and new directions. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New product diffusion models* (Chapter 7). Boston: Kluwer Academic Publishers.

Berenson, M. L., Levine, D. M., & Krehbiel, T. C. (2001). *Basic business statistics: concepts and applications* (8th ed.). New Jersey: Prentice Hall.

Boyce, W. E. & DiPrima, R. C. (1977). *Elementary differential equations and boundary value problems* (3rd ed.). New York: John Wiley & Sons, 112.

Chaterjee, R., et al. (2000). Dynamic models incorporating competition. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New product diffusion models* (Chapter 8). Boston: Kluwer Academic Publishers.

Ettlie, J. E. (n.d.). *Managing new process and product development*. RIT: MBA course.

Holt, C. (n.d.). *The history of docuTech*. RIT: presentation.

Lilien, G. L., et al. (2000). Diffusion models: Managerial applications and software. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New product diffusion models* (Chapter 12). Boston: Kluwer Academic Publishers.

Mahajan, V., Muller, E., & Wind, Y. (Eds.). (2000). *New product diffusion models*. Boston: Kluwer Academic Publishers.

Ratchford, B. T., et al. (2000). Diffusion models with replacement purchases. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New product diffusion models* (Chapter 6). Boston: Kluwer Academic Publishers.

Roberts, J. H., et al. (2000). Disaggregate-level diffusion models. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New product diffusion models* (Chapter 9). Boston: Kluwer Academic Publishers.

TrendWatch Graphic Arts. (2003, May) *Digital color printing: Ready for prime time*. New York: Author.

U.S. Census Bureau. (n.d.). *Industry Statistics Sampler: NAICS 323 Printing and related support activities* [data files]. Retrieved from http://www.census.gov/epcd/ec97/industry/E323.HTM

Appendix A: Solution to the Bass New Product Diffusion Equation with Time-Dependent Diffusion Parameters

The differential equation that describes the diffusion of new product, assuming only two possible customer states *s0* (has never adopted the product) or *s*2 (has one product from the primary market) is given by the following (see Equation 9):

$$\frac{dN}{dt} = \frac{1}{m}\frac{dm}{dt}N + P(a2]t, t+d]|\,s0)*$$
$$(m(t) - N(t)) \quad . \qquad (25)$$

Assuming Equation 10 for P(a2]t,t+dt]|s0), this becomes:

$$\frac{dN}{dt} = p(t)m(t) + (\frac{1}{m}\frac{dm}{dt} + q(t) -$$
$$p(t))N(t) - \frac{q(t)}{m(t)}N(t)^2 \quad . \qquad (26)$$

This equation is of a general form, known as the Ricatti differential equation:

$$\frac{dw(t)}{dt} = q_o(t) + q_1(t)w(t) + q_2(t)w(t)^2 \quad ,$$

which can be reduced to a linear differential equation by the substitution

$$w(t) = -\frac{1}{q_2(t)Y(t)}\frac{dY}{dt} \quad .$$

Applied to Equation 26 this translates into:

$$N(t) = \frac{m(t)}{q(t)}\frac{1}{Y(y)}\frac{dY}{dt} \quad , \qquad (27)$$

which, after substitution into Equation 26 leads to

$$\frac{d^2Y}{dt^2} + \left[(p-q) - \frac{dq}{dt}\right]\frac{dY}{dt} - pqY = 0 \quad , \qquad (28)$$

which is a linear homogeneous differential equation and which does not contain the market size anymore, only the coefficients *p* and *q*. The Equation 28 is easier to solve than Equation 26, and the exact solution of Equation 28 will depend on the exact time dependence of *p* and *q*.

If *p* and *q* are time-independent, then Equation 28 reduces to a linear homogeneous equation with constant coefficients, and the general solution is given by:

$$Y(t) = C_1 e^{z_1 t} + C_2 e^{z_2 t} \quad , \qquad (29)$$

where z_1 and z_2 are solutions of the quadratic equation:

$$z^2 + (p-q)z - pq = 0 \qquad (30)$$

which are given by z_1=q and z_2= -p.

The solution to the Bass equation, with time dependent market size *m*(*t*) and time-independent diffusion parameters *p* and *q* is therefore given by:

$$N(t) = \frac{m(t)}{q}\frac{C_1 z_1 e^{z_1 t} + C_2 z_2 e^{z_2 t}}{C_1 e^{z_1 t} + C_2 e^{z_2 t}} \quad . \qquad (31)$$

Combined with the boundary condition
N(0)=0, and thus:

$$C_2 = -C_1 \frac{z_1}{z_2} ,$$

Equation 31 finally becomes:

$$N(t) = m(t) \frac{e^{qt} - e^{-pt}}{e^{qt} + \frac{q}{p} e^{-pt}} = m(t) \frac{1 - e^{(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}} \quad . \quad (32)$$

Appendix B: Discrete Versus Continuous Models

The Bass equation assumes a continuous sales process that is very close to the real-world marketing process, which consists of a set of activities promoting the product on a daily or even hourly basis and which on a yearly scale can be seen as a quasi-continuous process.

The cumulative sales, as given by Equation 23, represent the cumulative sales on a finer than yearly time scale, in its extreme on an hourly or even finer time scale. This justifies the use of a continuous model rather than a discrete model.

Sub-sampling (Equation 23) on a yearly time scale is more practical, because sales data is more readily available on a yearly basis. Moreover, such an approach also avoids possible seasonal effects within a year. Also, the determination of the model parameters p, q, m can be done using regular minimization techniques, such as least-square approaches.

Alternatively, a process can be used for which actual sales data is collected on a finer time scale and modeled as a continuous time function $S(u)$. Then the parameters p, q, and m could be determined by minimizing a continuous error function:

$$\varepsilon = \int_{o}^{t} (N(u) - S(u))^2 \, du \quad , \qquad (34)$$

which becomes minimal for p, q, and m combinations, which satisfy:

$$\frac{\partial \varepsilon}{\partial p} = \int_0^t 2(N(u) - S(u)) \frac{\partial N}{\partial p} = 0$$

$$\frac{\partial \varepsilon}{\partial q} = \int_0^t 2(N(u) - S(u)) \frac{\partial N}{\partial q} = 0$$

$$\frac{\partial \varepsilon}{\partial m} = \int_0^t 2(N(u) - S(u)) \frac{\partial N}{\partial m} = 0 \quad . \qquad (35)$$

Clearly this set of equations is more complicated to solve than the set of equations obtained by minimizing the discrete error function given by Equation 23. However, it is possible that the sensitivity of p, q, and m with respect to the time t is different from the sensitivity of p, q, and m with respect to the discrete time variable k used in Equation 23.

An even more sophisticated methodology would be to consider p and q as time-dependent parameters, while minimizing Equation 34, finding $N(t)$ from Equations 27 and 28. Such a method relates to variational calculus,[3] and is beyond the scope of this independent study.

Appendix C: Parameter Extraction Minimization Examples

As explained earlier, the extraction of the diffusion parameters p, q, and m is best done by means of some minimization technique. In this paper the Excel Solver was used in a number of different cases in order to find the parameters p, q, and m when different numbers of years of sales data are available.

Table 10 gives some examples of the final values that were found for p, q and m starting from different sets of initial values for these parameters, and for 3, 6, and 11 years of available sales data.

As can be seen from this table, the market size found by the minimization process follows the initial guess for the market size quite well for 3 years of available sales data, less well for 6 years of available sales data, and for 11 years of available sales data the market size almost consistently converged to the same number, close to 39,000 units. As mentioned earlier, internal organizational issues may have affected the sales data and conclusions around market size derived from this data may therefore be invalid.

Table 10 also indicates the value of the error functions E_3, E_6, and E_{11}, as defined by Equation 24. E_3 was used to find the parameters p, q, and m when 3 years of sales data are available, E_6 when six years of data are available, and E_{11} when 11 years of sales data are available.

The concept and difficulty of local minima trapping encountered during the parameter determination is illustrated in more detail in the following graphs.

Figure 20 shows the error function E_6 for different market sizes m and fixed p=0.0142 and q=0.362. The error function has a clear minimum for a market size of around 38,833, corresponding to the value in bold type given in Table 10.

Figure 21 shows the error function E_6 for a constant market size m of 38,833 and a constant value p of 0.0142 and yet for a variable parameter q. As can be seen, the error function values become minimal for q close to 0.362, which corresponds to the value in bold type given in Table 10.

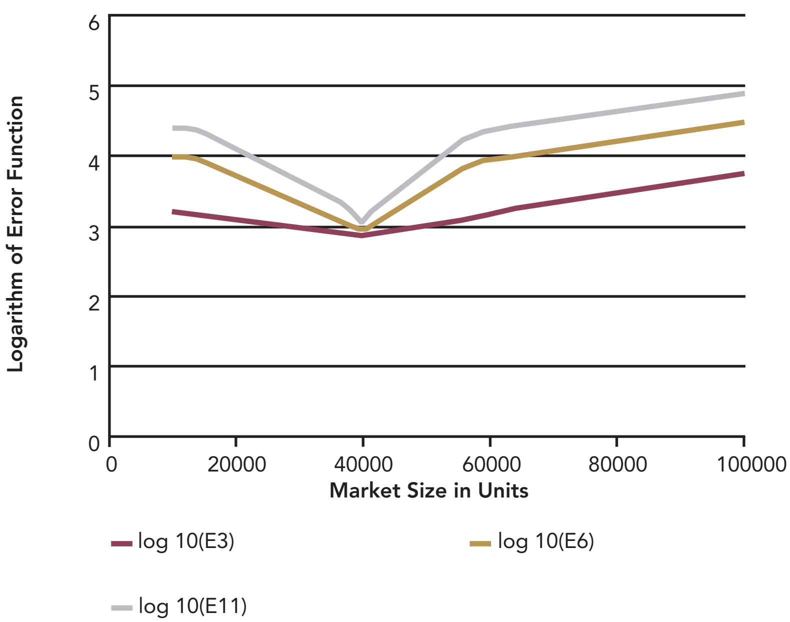

Figure 20. Error Functions E3, E6 and E11 for Fixed p (0.0142) and q (0.362), and Variable Market Size

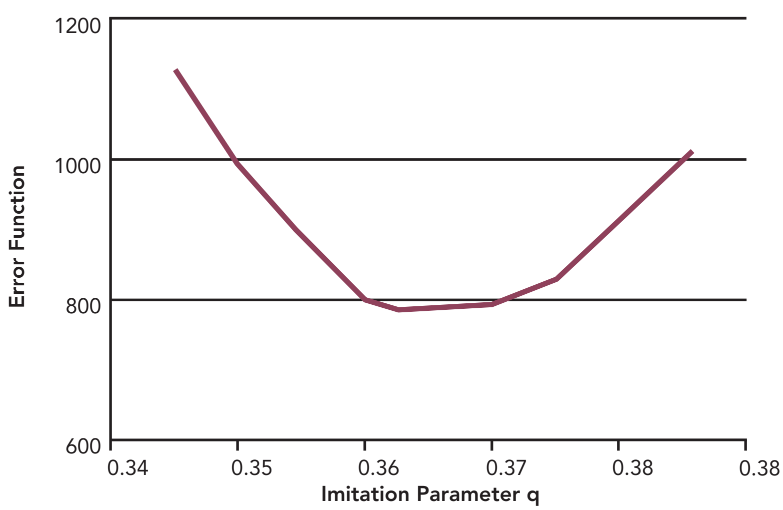

Figure 21. Error Function E6, for Fixed p=0.0142 and Fixed m=38833

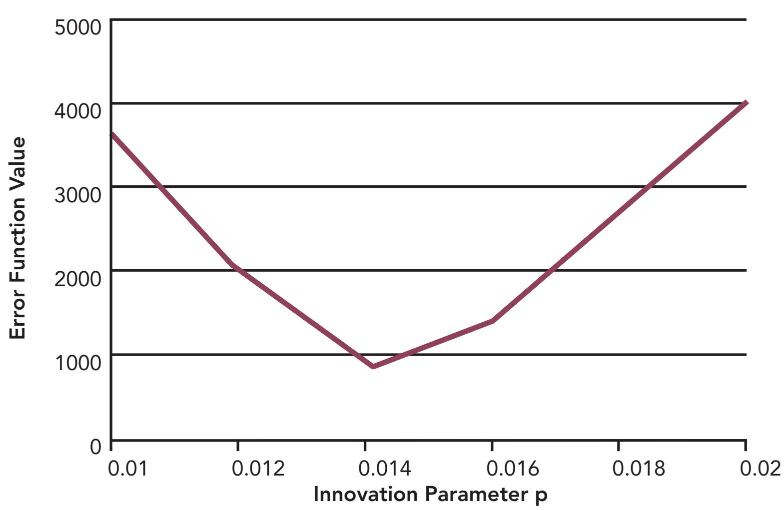

Figure 22. Error Function E6, for Fixed m=38833, Fixed q=0.362

Finally, Figure 22 shows the error function E_6 for fixed q=0.362 and fixed m=38,833 as a function of a variable factor of innovation p. As can be seen the error function reaches a minimum for a value of p very close to p=0.0142, corresponding to the value in bold type given in Table 10.

Combining Figures 20, 21, and 22 leads to the conclusion that the error function E_6 is minimized for the values (p, q, m)=(0.0142, 0.362, 38,833). These correspond to the bolded values in bold type given in Table 10. Such a minimum is referred to as a local minimum of the error function E_6.

When different initial guesses for p, q, and m are used to start the minimization process, then many different local minima are found that minimize the error function E_6 in the same way as illustrated by Figure 20, 21, and 22.

Figure 23 shows the error functions E_6 and E_{11} as functions of market size m for a number of values of m and their corresponding values of p and q minimizing E_6 in a local sense, in the same way as illustrated by Figures 20 to 22

As can be seen from Figure 22 the error function E_6, reaches an absolute minimum for values of p, q, and m of p=0.0228, q=0.725, m=13000, which are shown in italic in Table 10. One could argue that the latter values p, q, and m are therefore the better choice, since they minimize E_6 in an absolute sense. However, this is oversimplifying the problem, and the next paragraph elaborates on a better alternative.

If this processes of searching for an absolute minimum is repeated for different years, the resulting parameters p, q, and m will be considerably different, hence time-dependent. This contradicts the basic assumption of the Bass model and is therefore not a valid choice.

Figure 23 also shows the value of E_{11}, for the values of p, q, and m that minimize E_6. In other words, Figure 23 shows the values of E_{11} that are obtained by the following process:

- Using the first six years of sales data,
- Assuming an initial guess for the market size m,
- Calculating the p, q, and m parameters that minimize E_6 for that market size,
- Using the values of p, q, and m obtained in the previous steps to make the forecast and calculate E_{11}.

Figure 22 shows that for a market size of close to 38,800 units and values of p and q that minimize E_6 in a local way, the error function E_{11} calculated using the previous process reaches a clear minimum. This indicates that the same set of parameters p, q, and m can minimize the all the error functions E_k, for different values of k. This means that these values of p, q, and m are time independent and minimize all the error functions E_k in a local way. This set of p, q, and m parameters is therefore a better set, since it supports the fundamental assumption of the Bass model that the parameters are time independent.

Figure 23 also shows that the error function E_6 is quite insensitive to variations in m, once the market size is larger than 38,000 units. This means that a set of p and q values can be found that will lead to a similar forecast error for different estimated market sizes.

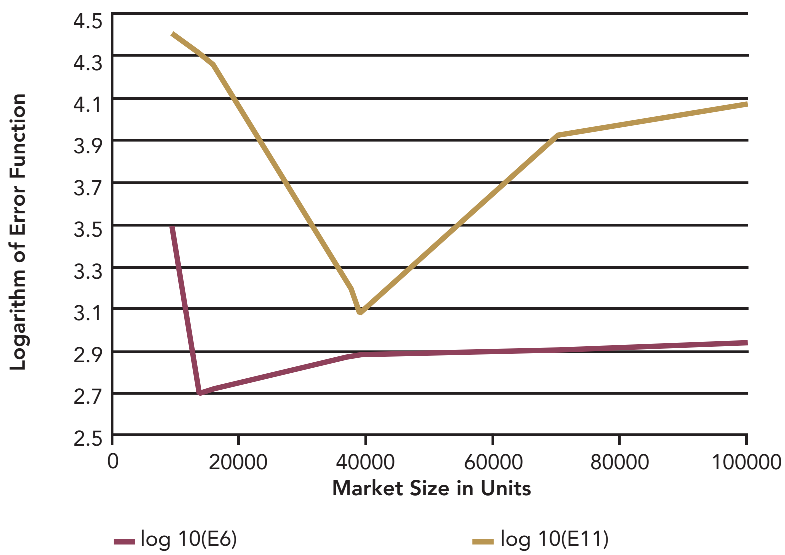

Figure 23. Error Functions E6, and E11, as a Function of Market Size, for Corresponding p and q Values that Minimize E6

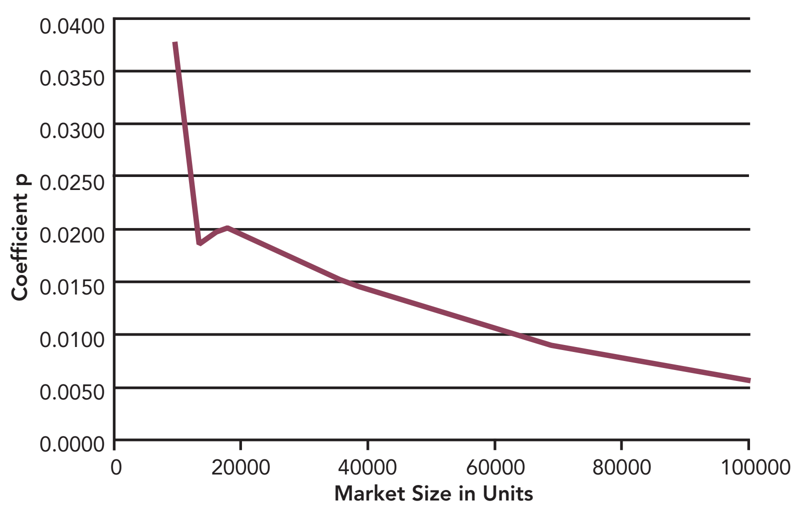

Figure 24. Factor of Innovation p that Minimizes E6 as a Function of a Given Market Size

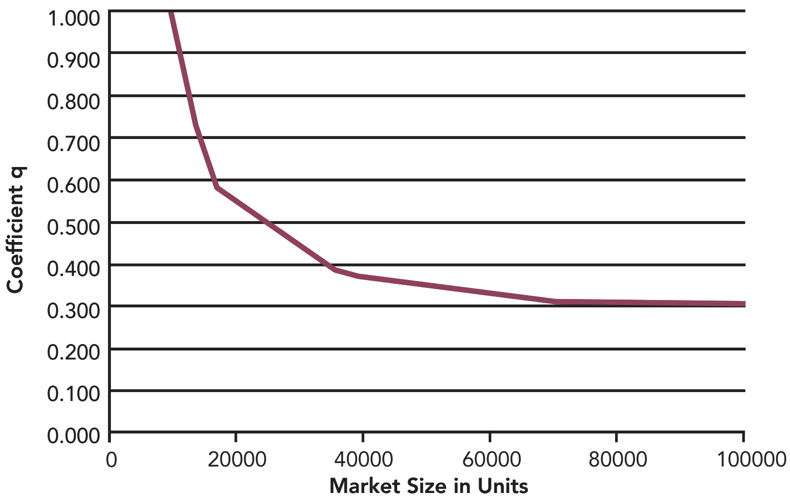

Figure 25. Factor of Imitation q that Minimizes E6, Given the Market Size

It is worthwhile to look at how the parameters *p* and *q* change for the different market sizes *m*. Figures 24 and 25 show the values of *p* and *q* that minimize E_6 for a given market size. It can be seen from these figures that the values of *p* and *q* depend substantially on *m*. As the assumed market size is increases, *p* and *q* are decreases. For very large market sizes, *p* asymptotically approaches 0.0012 and *q* approaches the value 0.295.

Initial Guesss			Three Years of Available Data – Final Solution					Six Years of Available Data – Final Solution						Eleven Years of Available Data – Final Solution				
p	q	m	p	q	m	E_3	E_{11}	p	q	m	E_3	E_6	E_{11}	p	q	m	E_3	E_{11}
0.0017	0.966	500000	0.0010	0.413	500000	659	83534	0.0012	0.295	5E+05	764	874	23333	0.0033	0.167	281777	1555	2306
0.0024	0.980	100000	0.0017	0.980	100000	219	153736	0.0059	0.306	1E+05	766	875	12167	0.0151	0.343	38834	771	889
0.0090	0.571	70000	0.0025	0.958	70000	217	98859	0.0084	0.318	70000	759	857	8701	0.0150	0.343	38850	769	889
0.0090	0.571	50000	0.0035	0.967	50000	215	62708	0.0246	0.587	15623	523	529	15250	0.0150	0.343	38839	768	889
0.0090	0.571	38834	0.0044	1.000	38834	213	42685	0.0245	0.620	14782	501	506	16344	0.0150	0.343	38832	767	889
0.0044	1.000	36000	0.0046	1.000	36000	213	37321	0.0147	0.385	35996	684	763	1755	0.0150	0.343	38834	768	889
0.0090	1.000	22000	0.0080	1.000	22000	211	14000	0.0246	0.587	15629	524	530	15242	0.0150	0.343	38832	768	889
0.0090	1.000	10000	0.0210	1.000	10000	261	24061	0.0370	1.000	10000	1696	3181	24192	0.0150	0.343	38832	768	889
0.0374	0.500	5000	0.0656	1.000	5000	600	34630	*0.0228*	*0.725*	*13001*	451	*495*	18883	0.0150	0.346	38598	761	889
0.0115	0.450	38833						**0.0142**	**0.362**	**38833**	593	**803**	1207					

p is the coefficient of innovation in the Bass diffusion model.
q is the coefficient of imitation in the Bass diffusion model.
m is the market size.
E_3, E_6, E_{11}, are error functions as defined by Equation 24. They are calculated for the values of *p*, *q*, and *m* found by the minimization process that minimizes E_3, if 3 years of sales data are available, E_6, if 6 years of sales data are available, and E11 if 11 years of data are available.

Table 10. *p*, *q*, *m* as a Function of Initial Guess and Number of Years of Available Sales Data

Appendix D: Digital Color Press Installation 1993-2003

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	Total 1993-2002	Estimated 2003 Total
Canon 31-50 ppm				150	2,100	3,600	4,500	5,700	6,700	7,000	6,000	29,750	35,750
Heidelberg 70 ppm									20	240	390	260	650
HP (Indigo)	20	167	300	180	189	195	215	280	310	360	410	2,216	2,626
Xeikon (all webs)	11	128	206	320	350	380	360	240	150	65	110	2,210	2,320
Xerox 30-40 ppm			120	900	1,600	1,900	1,100	600	0	0	200	6,220	6,420
Xerox 45-60 ppm								1,500	2,000	3,000	4,000	6,500	10,500
Xerox 100 ppm										65	120	65	185
Other									40	180	220	220	440
Total	**31**	**295**	**626**	**1,550**	**4,239**	**6,075**	**6,175**	**8,320**	**9,220**	**10,910**	**11,450**	**47,441**	58,891

The following numbers were obtained from Frank Romano at RIT and were published in *TrendWatch Graphic Arts*, May 2003.
The 2003 numbers were estimated by Frank Romano.

R·I·T Printing Industry Center
An Alfred P. Sloan Foundation Center

Rochester Institute of Technology
College of Imaging Arts and Sciences
55 Lomb Memorial Drive
Rochester, NY 14623
Phone: (585) 475-2733
http://print.rit.edu