"AGGREGATE DIFFUSION FORECASTING MODELS IN MARKETING: A CRITICAL REVIEW"

by

P. PARKER*

93/92/TM/MKT

* Associate Professor at INSEAD, Boulevard de Constance, Fontainebleau 77305 Cedex, France.

A working paper in the INSEAD Working Paper Series is intended as a means whereby a faculty researcher's thoughts and findings may be communicated to interested readers. The paper should be considered preliminary in nature and may require revision.

Printed at INSEAD, Fontainebleau, France
Aggregate Diffusion Forecasting Models in Marketing: A Critical Review

December 1993

Philip M. Parker
Associate Professor
INSEAD
77305 Fontainebleau France
(tel. 33-1-60-72-4000)

The author would like to thank Marcel Corsjens, Vijay Mahajan, Theodore Modis and Hermann Simon for their directive comments on earlier versions of this paper.
Aggregate Diffusion Forecasting Models in Marketing: A Critical Review

Abstract

Since the 1960s, a number of new product diffusion models have been developed and applied in marketing. This paper reviews the theoretical origins, specifications, data requirements, estimation procedures, and pre-launch calibration possibilities for these aggregate models. Following a critical review of both the problems and the potential benefits of these models, a number of suggestions are made with respect to future academic and applied research involving new product diffusion forecasting.
Aggregate Diffusion Forecasting Models in Marketing: A Critical Review

1. Introduction

One of the most important goals of firms is to successfully launch and market new products. Prior to launch or during the introductory months or years, market demand forecasts are critical to production, distribution, marketing and general planning efforts for these products. Since the 1960s, a number of new product forecasting models have been developed and documented in the academic marketing literature which are specifically designed to forecast new product acceptance over time. This paper reviews aggregate marketing models which find their theoretical roots in the literature on the diffusion of innovations (Rogers 1983). Diffusion processes characterize the acceptance or penetration of a new idea, behavior or physical innovation over time by a given social system. Based on hundreds of studies in various disciplines (sociology, education, marketing, economics), Rogers (1983) proposes that diffusion rates are affected by the type of innovation, its perceived attributes, the influence of communication channels, the nature of the social system and its population, the marketing activities of firms (change agents), consumer adoption processes, and the characteristics of individuals within the social system (personality, demographics, communication behavior). Robertson (1967), among others, extended the diffusion literature to marketing by suggesting that new product innovations undergo diffusion processes. Despite the number of factors that might affect new product acceptance, there has been a proliferation of parsimonious mathematical models over the last three decades which have been designed to characterize certain regularities in aggregate diffusion patterns, product life cycle curves, or new product penetration curves -- starting with the bell-shaped adoption curve, or the s-shaped penetration curve, as shown in Exhibit 1 for twelve consumer durable products.¹

¹ The data shown are first purchases only (total sales, less repeat and multiple purchases); source: various issues of Merchandising, Merchandising Weekly, Electronics Weekly, Dealerscope, Dealership Merchandising, Appliance and the Electronics Industry Association and A.C. Nielsen (see Parker 1992 and 1993a for details).
This review evaluates the use of these models as applied tools to forecast new product diffusion. Despite a number of documented industrial applications (see, for example, Bass 1986), a number of authors have raised concerns over the use of these models for forecasting purposes (Bernhardt and MacKensie 1972; Heeler and Hustad 1980). Mahajan, Muller and Bass (1990, p. 9) note that "by the time sufficient observations have developed for reliable estimation, it is too late to use the estimates for forecasting purposes." Simon (1992, p. 13) recently concludes that "as forecasting tools" diffusion models "are risky and potentially misleading." These concerns are generated, in part, by a fundamental lack of common knowledge concerning these models. Given the sheer number and variety of specifications, both applied and academic researchers seeking to use these models in a forecasting exercise are faced with the following questions:

- which model among the various alternatives should be used? For which types of products and marketing processes are the models correctly specified?

- lacking sufficient historical data, how can the models be used to predict future diffusion patterns?

- given some historical data, which estimation procedure is appropriate?

- how much does forecasting accuracy, or parameter stability improve as more observations become available? How many degrees of freedom are typically required before these models become useful (i.e. better than alternative methods)?

While a number of authors have suggested various answers to these questions, general and systematic approaches to using new product diffusion models have yet to be proposed in the literature. This review will attempt to clarify many of these issues and, where a general understanding is lacking, suggest future research directions. Our discussion is limited, however, to forecasting models and issues alone, given the richness of this area.

A number of previous literature reviews consider in greater detail the descriptive, theoretical
and normative use of diffusion models in marketing. Mahajan, Muller and Bass (1990), Bridges, Coughlan and Kalish (1991), and Lilien, Kotler and Moorthy (1992) provide excellent reviews of the entire modelling area. These supersede earlier reviews by Mahajan and Muller (1979), Meade (1984) and Mahajan and Wind (1986). Jeuland (1981) traces parsimonious diffusion models to their origins in a historical review of diffusion models across disciplines. The use of diffusion models in normative studies, in addition to the above, is reviewed in Hanssens, Parsons, and Schultz (1992, Chapter 8), and Simon (1989, Chapter 6). Since this review will focus on aggregate adoption (first purchase) models, we will not consider a number of new product forecasting procedures designed for frequently purchased products for which the trial process is not typically modelled as a diffusion process or where repeat purchase sales are emphasized; the reader is referred to the comprehensive reviews provided in Lilien, Kotler and Moorthy (1992, pp. 480-500) and Mahajan and Wind (1988). Finally, this review is limited to considering alternatives proposed within this particular literature stream; the relative merits of other approaches over diffusion models are only discussed if such comparisons are previously reported in the literature.

2. The Basic Theory

Unlike alternative forecasting methods, diffusion models have a common foundation in the diffusion of innovations literature. Rogers (1983, p.10) concisely defines diffusion "as the process by which an innovation is communicated through certain channels over time among members of the social system." Rogers identifies the four main elements of diffusion as the innovation, communication channels, time, and the social system. Innovations are typically accepted within a given social system by being adopted over time by various user segments: "innovators" (the first 2.5% to adopt), then the early adopters, the early majority, the late majority, and the laggards (the last 16% to adopt). Adoption levels are initially low by innovators and early adopters who act as opinion leaders, increase to a peak based on social learning, word-of-mouth and other forms of interpersonal influence, and then decline as the social system saturates; penetration levels follow the familiar s-shape pattern reflecting cumulative adoptions over time, and the bell-shaped diffusion curve for new adoptions (Exhibits 1 and 2 respectively). Forecasters are critically interested in having an accurate
picture of these two curves prior to, or in the early years following product introduction.

Extending diffusion research to consumer behavior, Robertson (1969, 1971) proposes that new products follow similar diffusion patterns. Robertson defines three types of innovations which undergo diffusion processes: continuous, dynamically continuous, and discontinuous innovations. In theory, there is a negative relationship between the continuity of an innovation and the level of social learning or interpersonal influence (including opinion leadership) among adopters; there is a negative relationship between social learning and diffusion speed (Gatignon and Robertson 1985). For a given market potential (target customer group), a continuous innovation will not disrupt established behavioral patterns after it is introduced (e.g., "new improved" dishwasher detergent), and is most likely to have an exponential penetration curve (Gatignon and Robertson 1985). A dynamically continuous innovation will cause some changes in behavior but not a significant amount within the market potential targeted (e.g., cordless telephones). Discontinuous innovations cause significant changes in consumer behavior and require substantial learning prior to adoption for members of the target market (e.g., the first television sets). Robertson suggests that dynamically continuous or discontinuous product innovations are likely to have bell-shaped diffusion patterns or s-shaped penetration curves similar to those found by Rogers. It is from this basis that new product diffusion models find their theoretical justification in marketing. In general, aggregate diffusion models have two basic components (1) a target market potential, or number of ultimate adopters, and (2) a growth rate component which can characterize interpersonal influence among members of the target market. Models lacking these components include naïve extrapolation and various time-dependent growth models. We begin our review by highlighting various model Specifications which, by capturing these components, have proven useful to forecast the size of an innovation’s market potential, and its speed of adoption.
3. Choosing Among the Proposed Models

Despite the concerns raised in the introduction, successful applications of diffusion models as applied forecasting tools have been documented (see, for example, Mahajan, Muller and Bass 1990). The key to their successful use depends, in part, on the researcher correctly matching the model to the adoption process under study. To guide this choice, we will briefly describe three types of aggregate diffusion models: (1) first purchase models which stem from or are generalized by the model proposed by Bass (1969), (2) non-Bass first purchase models, and (3) repeat purchase models. First purchase models generally multiply a hazard rate in each time period against an aggregate market potential. The hazard approach is the general framework for most diffusion models in marketing. We will also briefly review recent efforts which relax certain assumptions common in these aggregate diffusion models, namely that the probabilities of adoption are homogeneous across individuals (see, for example, Chatterjee and Eliashberg 1990). Micro-level diffusion models consider individual-level adoption processes and have been found superior to certain aggregate hazard-based models in forecasting new product adoption (see, for example, Lattin and Roberts 1989, and Sinha and Chandrashekaran 1992). Finally, repeat purchase models capture additional purchases (adoptions) after first purchases. In this section we primarily focus on reviewing the functional forms of these models, leaving data requirements, estimation, and forecasting issues to later sections.

3.1 First Purchase Diffusion Models

Fourt and Woodlock (1960) proposed one of the earliest new product growth models in marketing which has seen wide industrial application. If f(t) is the density function of a consumer to adopt a product in time t and F(t) is its associated cumulative distribution function, Fourt and Woodlock assume that the conditional probability of adopting, given that one has not yet adopted, is constant over time:

\[ \frac{f(t)}{1-F(t)} = a \]  
(1)
or,

\[ f(t) = a (1 - F(t)) \]  \hspace{1cm} (2)

where \( a \) is a positive hazard rate. To forecast trials (first adoption/purchases), one can define \( M \) as the market potential, or the total number of ultimate adopters (assuming this is static), \( n(t) \) as the number of adopters in time \( t \), and \( N(t) \) as the total number of cumulative adopters up to, but not including \( t \). Since \( f(t) = n(t)/M \), and \( F(t) = N(t)/M \), then,

\[ n(t) = a (M - N(t)) \]  \hspace{1cm} (3)

growth saturation effects effects

Typical of aggregate diffusion models in marketing, this model exhibits growth effects (the hazard rate) which regulate the speed and shape of the diffusion process, and saturation effects which constrain the process to a fixed level \( M \). Consistent with Gatignon and Robertson (1985), the Fourt and Woodlock model has typically been applied to the trials or first purchases of inexpensive packaged products, continuous innovations which have exponential trial curves or products having low word-of-mouth effects (e.g., the PROMPT system, developed by Information Resources Inc., applies the model to scanner data to generate trial forecasts).

Not all products experience exponential penetration rates beginning at launch. The model proposed by Bass (1969) generates S-shaped penetration curves while explicitly considering the diffusion literature. Departing from Rogers, Bass defines two categories of adopters who interact with each other in a contagion process: innovators and imitators. The Bass model postulates that the hazard rate in time \( t \) is a linear function which increases with the cumulative distribution function \( (a + b F(t)) \), resulting in the following aggregate diffusion model:

\[ f(t) = \frac{d}{dt} \left( a + b F(t) \right) \]

\footnote{Fourt and Woodlock note that \( M \) can represent some percent of a given social system; for example, \( M \) might represent 40 percent of a target market.}
\[
n(t) = (a + b \frac{N(t)}{M}) (M - N(t))
\]

where \(a\) and \(b\) are positive constants, labelled the coefficients of "innovation" and "imitation" respectively, \(n(t)\) are first purchases in time \(t\), and \(N(t)\) are first purchases up to but not including \(t\). Equation 4 presents the core structure of most parsimonious aggregate diffusion models suggested in the marketing literature. The Bass model enjoyed early acceptance among modellers for a number of reasons. First, it generalized both the Fourt and Woodlock (1960) model, which sets \(b=0\), and the Mansfield (1961) model or simple logistic penetration curve, which sets \(a=0\); this later model has been successfully applied to a number of categories exhibiting high word-of-mouth or diffusion effects (see Chadda and Chitgopekar 1971; Modis and DeBecker 1988, 1992; Olshavsky 1980; and Modis 1993, 1994 for various applications). Second, the Bass model had early successes in fitting diffusion processes for a number of industries: consumer durable goods, retailing, agriculture, industrial processes, medical equipment and telecommunications (Akinola 1986; Bass 1969; Dodds 1973; Nevers 1972). Third, the model represents an appealing alternative to simple extrapolation or time series techniques which would typically fail to predict the peak in first purchase sales. While recent methods have been proposed to forecast time series which have pattern changes beyond the historical data, these would rely on the number and extent of similar changes in historical data (see, for example, Carbone and Makridakis, 1986). In typical new product forecasting applications, however, such historical changes are absent (i.e. the first purchase sales peak is unique).

Two major issues have been voiced concerning the basic Bass formulation. First, some authors question the interpretation of the parameters as representing the proportion of "innovators" and "imitators" in the social system, since the respective coefficients exert influence throughout the diffusion process, or the influence/information flows are similar across all adopters (Tanny and Derzko 1988). Lekvall and Wahlbin (1973) and Mahajan, Muller and Srivastava (1990), for example, propose labelling \(a\) the coefficient of external
influence (capturing information external to cumulative social experience) and $b$ the coefficient of internal influence. To date, few studies have explicitly tested whether, for example, products having larger diffusion parameters values actually exhibited more word-of-mouth influence (or other), versus products with small values. Since diffusion parameters reflect the net effect of numerous phenomena, it is not clear that they can be generally labelled as reflecting certain influences, especially in a general sense across all categories. Second, and despite the apparent advantages over alternatives, a number of authors note that the model fails to consistently predict either the peak in new product sales or the penetration ceiling for a broad range of categories (Bernhardt and MacKensie 1972; Heeler and Hustad 1980; and Tigert and Farivar 1981), and that direct parameter estimates are often implausible or provide nonsensical projections. This problem is especially critical when external judgements are not imposed on certain model parameters (e.g. the market potential, M). Exhibit 2, for example, depicts the first purchase diffusion curves, or adoption life cycles for various consumer durable products. Unlike the theoretical adoption pattern proposed by Rogers, and characterized by the Bass model, some categories are skewed left, skewed right, or have multiple peaks. Lacking external constraints, the Bass model will inadvertently interpret early variations in adoption (e.g. due to seasonality, business cycles, data reporting errors, etc.) as inflection points, or peaks, which can lead to unreasonable parameter estimates and long-run forecasts. During the late 1970s and 1980s a variety of extensions to the original Bass model were proposed to overcome some of these shortcomings, including the addition of various marketing mix elements, and an extensive study of parameter estimation procedures.

Modelling extensions have incorporated dynamic market potentials (Mahajan and Peterson 1978, Mahajan, Peterson, Jain and Malhotra 1979), non-uniform interpersonal influences (Easingwood, Mahajan and Muller 1983), and heterogeneous adopter populations (Jeuland 1981). The following 5-parameter diffusion model, interpreted below, simultaneously represents these modifications to the Bass model:
\[ n(t) = \left( a + b \left( \frac{N(t)}{cM(t)} \right)^{(1+d)} \right) (cM(t) - N(t))^{(1+e)} \]  

where \( a, b, c, d \) and \( e \) are estimated constants, and \( M(t) \) is the market potential (social system) which varies over time. Equation 5 belongs to a class of flexible diffusion models which allows for non-symmetric patterns (see Mahajan, Muller and Bass 1990). The imitation (or internal influence) coefficient, \( b \), is the baseline level of interpersonal influences (or contagion effects). The parameter \( c \) measures the long-run penetration ceiling and accounts for the non-adoption by some percentage of the market potential \( M(t) \). \( M(t) \) is set exogenously (say, the number of households wired with electricity when studying consumer electronics); see, for example, Schmittlein and Mahajan (1982), and Kamakura and Balasubramanian (1988). Values of \( c \) are positively related, therefore, to the number of segments that ultimately adopt the product within a social system or target market. Low empirical estimates of \( c \) indicate that the long-run diffusion process is limited to a small proportion of households -- perhaps one or two segments. A value of .99 for \( c \) indicates that a product is adopted by virtually all members in a given social system (e.g. as was the case for black-and-white television in the United States). Since influences may intensify, or diminish as the category matures, Easingwood, Mahajan and Muller (1981, 1983) control for nonuniform interpersonal influence (NUI) by incorporating the constant \( d \) (-1<\(d\)). Holding \( b \) constant, lower values of \( d \) result in faster penetration take-off periods (interpersonal influences increase), while larger values of \( d \) reflect increased resistance. The NUI parameter is analogous, therefore, to changes in consumers' immunity to social influence or contagion effects. The NUI parameter allows the adoption curve to be skewed left, or right, in a non-symmetric pattern. Finally, Jeuland (1981) controls for heterogeneous adopter populations via the parameter \( e \) (0<\(e\)). Since not all members of the social system are identical in their income, tastes, ages, etc., Jeuland suggests that \( e \) controls for differences across consumers in their propensities to adopt. Within equation 5 there are some 20-nested models which have all seen some empirical application in the marketing (see, for example, Easingwood et al. 1981, 1983; Kamakura and Balasubramanian 1988; Parker 1992, 1993a; and Schmittlein and Mahajan 1982).
A number of additional studies have modified the Bass model to consider more complex marketing issues including supply restrictions (Jain, Mahajan and Muller 1991), distribution and geographic diffusion (Mahajan and Peterson 1979; Jones and Ritz 1991), substitutes, complements and successive product generations (Peterson and Mahajan 1978; Norton and Bass 1987), brand-level diffusion (Parker and Gatignon 1994), cross cultural effects (Gatignon et al. 1990, Takada and Jain 1991), software piracy (Givon et al. 1993), sampling (Jain, Mahajan and Muller 1991), advertising (Horsky 1983, Simon and Sebastian 1987), pricing and income effects (Robinson and Lakhani 1975; Bass 1980; Rao and Bass 1985; Kamakura and Balasubramanian 1988; Horsky 1990; Jain and Rao 1990; and Parker 1992), price expectations (Narasimhan 1989), and other marketing decision variables (e.g. Bass and Krishnan 1992, Jain 1992, Kalish and Sen 1986). These studies generally are descriptive or use diffusion models to test research hypotheses (see also Mahajan, Muller and Bass 1990). While they provide valuable information on diffusion processes, many do not illustrate forecasting applications as the additional variables considered may prove difficult to forecast (prices, advertising, competitive actions, etc.). Parallel to these empirical descriptive studies are normative studies which derive optimal marketing mix strategies (see Mahajan, Muller and Bass 1990; Hanssens, Parsons, and Schultz 1992, Chapter 8; and Simon 1989, Chapter 6 for comprehensive reviews of this literature). These studies typically use general formulations to describe the diffusion process and do not, therefore, consider forecasting issues. Finally, some authors have proposed models which relax the deterministic nature of the parameters in aggregate diffusion models. Noting the poor fit of the original Bass model to certain categories, some authors have suggested that the parameters reflect stochastic processes (see for example, Lilien, Rao and Kalish 1981; and Eliashberg, Tapiero and Wind 1987). These approaches typically require a larger number of parameters without specifically modelling the factors which generate the uncertainty observed. Noting the limited use of such procedures in the literature, Bridges, Coughlan and Kalish (1991, p.265) conclude that "little incremental analytical insight is currently available due to taking a stochastic approach to modelling diffusion."

In addition to Bass-type diffusion models, any number of growth models can be used to characterize cumulative distribution functions, and, therefore, s-shaped penetration curves; see
for example, Narayanan (1992) who uses the Burr distribution to characterize diffusion processes, Sharif and Islam (1980) who use the Weibull distribution, Stapleton (1976) who proposes the normal distribution, or Trajtenberg and Yitzhaki (1989) who suggests a Gini’s expected mean difference approach. Models, which have both diffusion and saturation terms analogous to Bass-type models are summarized below (for additional models see the reviews in Mahajan, Muller and Bass 1990, Young 1993, and Young and Ord 1990):

\[
\begin{align*}
n(t) &= b \left( \frac{N(t)}{M(t)} \right) \left( \ln \frac{M}{N(t)} \right) & (6) \\
n(t) &= b \left( \frac{N(t)}{M(t)} \right) \left( M - N(t) \right)^2 & (7) \\
n(t) &= b \left( \frac{N(t)}{M(t)} \right) \left( M(t) - N(t) \right)^k & (8) \\
n(t) &= b/t \left( \frac{N(t)}{M(t)} \right) \left( M(t) - N(t) \right) & (9)
\end{align*}
\]

Equation 6 is the Gompertz curve (see Hendry 1972, Dixon 1980, Young and Ord 1985) and has been applied to consumer durables and agricultural innovations; equation 7 is the Floyd (1962) model and has been applied to industrial innovations; equation 8 is proposed by Nelder (1962) and has been applied to agricultural innovations (\(k\) is constant; see also McGowan 1986, Von Bertalanffy 1957, and Richards 1959 for similar formulations); equation 9 is proposed by the Stanford Research Institute (SRI; Teotia and Raju 1986) and has been applied to innovations in the energy sector. A variety of other growth models have also been proposed; see, for example, Bewley and Fiebig (1988), Blackman (1971), Bodger and Tay (1987), Brockhoff (1967), Fisher and Pry (1971), Gore and Lavaraj (1987), Harvey (1984), Kumar and Kumar (1992a, 1992b), Lancaster and Wright (1983), Lee and Lu (1987), Levenbach and Reuter (1976), Meade 1989, Pozanski (1983), Sharif and Kabir (1976), Sharif and Ramanatha (1981), Skiadas (1986), Steyer (1993), Tingyan (1990). In contrast to Bass-type models, these alternatives have seen little direct application in marketing and have not undergone modifications and extensions to the same extent as the original Bass model. Furthermore, in many cases these models do not directly relate to diffusion theory and have parameter values which do not lend themselves to external interpretation (other than a market saturation term, if this is given).

As the number of new models proposed has generally outpaced forecasting applications, there
has been numerous calls for systematic comparisons among the available models (Mahajan, Muller and Bass 1990, p. 21; Lilien, Kotler and Moorthy 1992, p. 480). A number of reasons justify such comparisons. First, since many diffusion models have been developed for normative and descriptive purposes, forecasting issues simply go unaddressed. Second, studies will often examine one particular issue in isolation, without necessarily considering or incorporating the incremental improvements identified or made in previous studies. The resulting "parallel processing" in the literature has lead to proposed models being compared to the original Bass (1969) model without direct comparison to flexible forms developed since, especially models which allow for skewed diffusion patterns (e.g., the non-uniform influence, NUI, parameter, \(d\)), or an underlying social system growth rate. Furthermore, proposed models often use the original Bass model as a point of departure, leading to the exclusion of parameters or phenomenon found generally relevant in characterizing diffusion processes. More importantly, many diffusion models have not been illustrated by or for practising econometricians or forecasters, leading to a less than ideal statistical treatment in most studies:

- studies are descriptive and typically illustrate "fit"; improvements in forecasts over alternative methods (which are not examined or identified) are unknown;

- basic fit and diagnostic statistics are frequently unreported, especially indicating collinearity and serial correlation; basic forecasting measures based on hold-out samples are often omitted (mean absolute deviations, mean absolute percentage errors, root mean square errors, Theil's, 1966, U statistic, etc.);

- certain statistics are incorrectly reported; for example, typical nonlinear diffusion models do not include a linear intercept term which prevents the traditional use of the classic R-squared statistics, yet these are reported as indications of fit;

- fit statistics are often not comparable across studies; for the same category one can calibrate the models against the penetration curve (\(N(t)\)) which is smoothed and will likely improve fit, versus the adoption curve (\(n(t)\)) which typically has greater variation and
lower fit statistics;

models are typically not compared using rigorous statistical tests (e.g. likelihood ratio tests, or non-nested nonlinear model tests); exceptions using diffusion data include Balasubramanian and Jain (1993), Parker (1992, 1993a) and Parker and Gatignon (1994).

Despite these shortcomings, some progress has been made to assist researchers in selecting from among the various models. In a meta-analytic and econometric study of long-run diffusion patterns from a sample of both highly successful and relatively unsuccessful durables product categories, Parker (1993a) reports that aggregate diffusion models essentially require no more than four parameters to adequately characterize ("fit") long-run diffusion patterns observed in practice. While considering models with various levels of parameterization (from one parameter to five parameters) using nested and non-nested likelihood ratio tests, the study finds the following model to be statistically superior among the 24 alternative models studied:

\[ n(t) = (a + b \frac{N(t)}{cM(t)^{(1+d)}}) (cM(t) - N(t)) \]  

(10)

diffusion effects   saturation effects

where \(a, b, c\) and \(d\) are estimated constants. Equation 10 adds to the NUI model proposed by Easingwood et al. (1983) a dynamic market potential (based on underlying population or social system growth) and \(c\) which measures the ultimate penetration level among social system members; \(M(t)\) is defined as the number of households, which varies over time. If parsimony is desired, the study finds that models nested within equation 10 fit better than models, including non-Bass models, which are not nested within equation 10, assuming an equal number of parameters across the models being compared. The best one, two and three parameter models nested within equation 10 contain \((b)\), \((b, d)\), and \((b, c, d)\) or \((a, b, d)\) respectively. The incremental fit (measured as a simple correlation between actual and fitted values), improves from .59 to .74 to .81 to .86 across the best fitting one, two, three and four parameter models respectively (these differences are statistically significant). Exhibit 3 illustrates the incremental effect of each parameter on aggregate diffusion curves, holding
others in the model constant. As described earlier, the external influence coefficient, $a$, does not affect the overall shape of the adoption curve, but does affect the intercept, and, therefore, the years needed to reach the peak. The market potential parameter, $c$, does not affect the years to peak, but the absolute magnitude of adoptions in any given year. The diffusion parameter $b$ affects both the timing and amplitude of the peak, and the NUI parameter reflects the skew of the adoption curve, and, therefore, the peak's timing. The 19 product categories shown in Exhibit 2 illustrate the model's ability to characterize a variety of diffusion patterns. It is interesting to note that in none of the 19 cases was the original Bass model the best fitting model, though, on average, it was equivalent to the best alternative three parameter model. Since equation 10 is the integration of several incremental improvements to the original Bass model and has only recently been empirically evaluated, there is no report of its use in normative studies, though models nested within equation 11 have been considered in descriptive studies of pricing and the diffusion process (Kamakura and Balasubramanian 1988; Parker 1992; Balasubramanian and Jain 1993).

Since the conclusion that equation 10 is statistically superior to alternatives is based on fit statistics, it is not clear that it will provide adequate forecasts. Research conducted by Rao (1985) evaluates the forecasting accuracy of six diffusion models and five naïve trend extrapolation models using data from four consumer durables categories. Rao finds that diffusion models generally outperform naïve models, and that among diffusion models, the following model, which represents the Bass model adjusted for nonuniform interpersonal influence (Easingwood et al., 1983), provides superior forecasts:

$$n(t) = (a + b \frac{N(t)}{M(t)}(1+d)) (M(t) - N(t))$$  \hspace{1cm} (11)$$

where $a$, $b$ and $d$ are estimated constants. In a second study comparing three diffusion models on six categories, Lavaraj and Gore (1990) find that there is no best long-run forecasting model. The study was apparently conducted without consideration of the findings of Rao (1985) and did not consider the model in equation 11. Lavaraj and Gore nevertheless find that the best forecasting model is typically the model which best fits the data.
Finally, Young (1993) compares the forecasting accuracy of nine growth models, including the original Bass (1969) model and a simplified model with non-uniform interpersonal influence proposed in Easingwood et al. (1981). Applying the models to a variety of data sets with varying lengths, Young (1993, p. 385) finds that the Bass model outperforms or equals non-Bass growth models, especially when diffusion is under the 50% takeover point. The simplified NUI model (excluding $a$ and $c$) was found less accurate when few observations are available for calibration, but was found best when the saturation level was externally imposed (known) and data sets contained more than 15 observations. Both the Bass and NUI models were found best in forecasting when the series had not reached 50 percent takeover level. The frequency with which the Bass and NUI models, combined, were deemed best was rivalled only by the logistic model proposed by Harvey (1984), which incorporates a time-dependent component in the diffusion process. While this comparison is encouraging, it is limited to short forecast horizons (1 year and 3 years), and internally estimated market potentials which are assumed to remained fixed over the estimation and forecast periods (often spanning multiple decades).

Combined, the existing comparative studies tend to indicate that the four-parameter formulation in equation 10, a modified NUI model, best fits long-run diffusion data. The model requires that the target market be externally identified and be allowed to vary over time ($M(t)$). One parameter reflects the initial penetration level ($a$), another the ultimate penetration ceiling ($c$), and the two remaining are required to characterize the shape or speed of the diffusion process ($b$ and $d$). If parsimony is desired, the next best three parameter model eliminates the NUI parameter ($d = 0$) or the original Bass model external influence coefficient ($a = 1$); see Parker (1993a). Models nested within equation 10, including the original Bass model and the NUI model proposed by Easingwood et al. 1983), generally provides superior (or equally good) forecasts compared to naïve or alternative diffusion model approaches. This conclusion, however, is limited to too few studies to be categorical. A more broad evaluation of the models proposed, using a far larger number of products in multiple categories and countries, is clearly needed to improve our general understanding.
3.2 Micro-level Diffusion Models

The models reviewed above treat a category at the aggregate level. In many marketing applications, especially for industrial markets, it is important to anticipate adoption timing at the individual level. Alternative approaches have been recently proposed which disaggregate the diffusion process to individual members of the social system (for more thorough reviews, see Mahajan, Muller and Bass 1990, and Sinha and Chandrashekaran 1992). While an aggregate diffusion model can forecast the aggregate number of first purchases in a given year (or time period), micro-level models can potentially forecast the adoption timing of each individual member of the social system, provided that sufficient data are available at that level. The aggregation of individual level adoption processes generates an aggregate diffusion curve. Since the unit of analysis is at the individual-level, various causal factors can be assumed to affect individual adoption timing. Factors previously considered include the distribution of knowledge, uncertainty, risk aversion, price/performance trade-offs, and consumer perceptions of product performance (Hiebert 1974; Stoneman 1981; Reinganum 1981; 1983; Feder and O’Mara 1982; Jensen 1982; McCardle 1985; Oren and Schwartz 1988; Lattin and Roberts 1989; Schmidt and Witte 1989; Chatterjee and Eliashberg 1990; Sinha and Chandrashekaran 1992). While pre-launch survey approaches have been suggested to estimate individual adoption which can be extrapolated to the aggregate level (Chatterjee and Eliashberg 1990), extensive empirical applications of these approaches have yet to be reported in the literature. This may result from the requirement to collect relevant data on consumer heterogeneity and individual-level adoption. Sinha and Chandrashekaran (1992) report the successful application of a split-hazard function approach to banks adopting automatic teller machines; a variety of variables are used to capture adopter heterogeneity: size, location, ownership by a holding company, local wage rates, and bank income. Future work in this area promises to be extremely useful in describing adoption processes and testing a variety of hypotheses concerning individual, or segment behavior (Weerahandi and Dalal 1992). In particular, the application of such approaches to international diffusion processes where the units of observation are countries would be a useful extension of this literature (i.e. at what moment will a particular country adopt a new technology or standard?). For now, insufficient experience is available to make generalizations on the usefulness of these approaches.
3.3 Repeat Purchase Extensions

Diffusion theory suggests that consumers undergo various stages when accepting and adopting an innovation. Rogers (1983, p. 164) suggests five separate stages: knowledge, persuasion, decision, implementation and confirmation. For the most part, diffusion models have typically characterized the first four stages, with implementation representing first purchases. A number of studies have considered the fifth stage which captures multiple purchases of the innovation after the initial adoption. In many cases the success of an innovation is not based on the initial adoption (or trial) level, but on the repeat purchases that follow. The models presented above are purely adoption (first purchase) oriented. Repeat purchase diffusion models generally have the following form:

\[ s(t) = n(t) + r(t) \]

where \( s(t) \) are total sales, \( n(t) \) are first purchases, and \( r(t) \) and repeat purchases in time period \( t \). Most models of repeat-purchase diffusion processes disaggregate the adopter population into distinct sub-segments, or transition states. Transition equations model the flow between states along a structured hierarchy (e.g., non-awareness, awareness, trial, repeat, re-repeat, etc.). Each state undergoes some form of diffusion, the speed of which may be affected by marketing mix variables (see, for example, Lilien, Rao and Kalish 1981; Mahajan, Wind and Sharma 1983; and Rao and Yamada 1988). Typically these models are created for a particular product category (pharmaceuticals and consumer durables), given that hierarchies and data availability vary from one category to another. In the case of consumer durable products, \( r(t) \) is often specified as being a function of product breakdown, scrapping rates generated by a probability distribution, replacement cycles, or multiple unit ownership of the innovation studied (see, for example, Lawrence and Lawton 1981; Olson and Choi 1985; Kamakura and Balasubramanian 1987; Norton and Bass 1987; Bayus 1987; Bayus 1988; Bayus, Hong and Labe 1989; Parker 1991). As these models all depend on the initial adoption process, the issues of forecasting directly depend on the ability of the first purchase model to forecast adoption penetration. Repeat sales forecasts are generated based on early repeat sales estimates. For the remainder of this review, we will focus our attention,
therefore, on the applied use of the aggregate first purchase models reviewed above.

4. Estimating Diffusion Models

Beyond model selection, estimating models prior to, or during the early years following a product's launch becomes important. The methods proposed largely depend on the data available and the formulation hypothesized. In the following section we will discuss, in chronological order, the tasks facing forecasters: pre-launch, then post-launch estimation.

4.1 Pre-Launch Forecasting

Since major investments and marketing decisions are typically made during the earliest phases of product launch, the value of forecasts may decline as the number of observations become available. One might argue that this problem will likely increase in the future as product life cycles become shorter, though recent evidence indicates that such is not the case (Bayus 1992). In any event, it is clear that for diffusion models to have wide applicability and use, pre-launch parameter estimation methods must be explored. For this reason a number of authors have considered estimation procedures that can be used in the absence of data. One of the first questions a researcher must ask, however, is whether diffusion models are applicable to the category in question. Given their theoretical origins, one can strongly argue that the models presented can apply to all products (consumer durables, high technologies, or consumer packaged goods) provided that the object of the forecast is limited to the adoption or trial process of sales (otherwise, repeat purchase extensions, reviewed above are necessary).

Once a model (e.g. equation 10) is selected, a number of methodologies have been suggested to obtain pre-launch estimates of diffusion model parameters. Broadly summarized, many authors suggest using historical analogies and managerial judgements to calibrate model parameters (Geurts and Reinmuth 1980; Lawrence and Geurts 1984; Lawrence 1980). If enough analogies are available, one can obtain estimates of parameters for the analogies and regress these estimates against various factors likely to affect the diffusion process. Product
forecasts are generated by characterizing new innovations for each of the factors identified (see for example, Srivastava et al. 1985; Gatignon et al., 1989; and Montgomery and Srinivasan 1989). In the absence of multiple analogies, one might simply use parameter estimates from a similar product as priors (Lilien, Rao and Kalish 1981). Again, if the goal of the forecast is the long-run adoption cycle, three aspects must be considered: the market potential and the long run penetration ceiling (asymptote) of the penetration curve, represented by coefficient c in equation 10; (2) the initial or first year penetration level which acts as the intercept of the penetration curve, represented by the coefficient of innovation or external influence, a; and (3) the speed of adoption between these two levels, represented by the coefficients b and d in equation 10.

Pre-launch estimates of the penetration ceiling, c, first requires a careful definition of the target market and a forecast of how this target will evolve over time. For many categories this can be a trivial exercise (e.g. for electronic consumer durables M(t) can be households wired with electricity which has a stable growth pattern). For others, such as cellular telephones, this becomes complicated as the market to be forecasted spans multiple markets, or segments. In such cases, either data need to be forecasted (and ultimately collected) by market or segment, or these need to be aggregated in a judicious manner. Once the target market is defined, and forecasted, we must then estimate the long run percent that will eventually adopt, c (multiplied against M(t)). Most authors suggest estimating the long run market potential based on expert judgement which can involve managers "questimating" the percent share a product will have vis-a-vis existing alternatives (e.g. using decision calculus or delphi approaches). These judgements can be coupled with industry surveys incorporating various economic factors, extensive analyses of long run competitive substitutes, or a combination of methods (see, for example, Tigert and Farivar 1981; Clarke and Soutar 1982; Twiss 1984; Hauser and Urban 1986; Teotia and Raju 1986; and Thomas 1987). Otherwise, if products are successive generations to earlier innovations, or are contingent to existing innovations, market potentials can be estimated from these existing products (see Bayus 1987; Norton and Bass 1987). Rogers (1983) suggests that the long-run acceptance of an innovation should be linked to the product's perceived attributes: relative advantage, perceived risk, observability, trialability, complexity, and communicability. Given the large number of
factors which would tend to affect these perceptions and, therefore, the long run penetration level of most products, Horsky (1990) notes that the market potential "varies considerably across product categories and generalizations about its overall value are unlikely." A lack of such a general understanding has led most authors to suggest informed judgement of managers coupled with available industry data or information on the long run diffusion patterns for similar products to estimate the market potential. Again, for some categories, this may be a trivial exercise (e.g. it would be safe to assume that less than 50 percent of adults will adopt products which are gender specific, or that attachments will be limited to the long-run penetration of the existing innovation). For others, this estimate can become pure speculation (e.g. long run cellular telephone penetration). Clearly, there is a strong need for researchers to rigorously explore why some products seem to reach high long-run (e.g. 30-year) penetration levels (e.g. above 90 percent), versus moderate or very low levels (below 5 percent). Such investigation is best performed on large data sets covering multiple categories and multiple countries to insure that generalizations are not context specific.

To estimate the initial penetration level, similar procedures to the ones suggested for the long-run market potential are typically recommended: industry studies/surveys or informed managerial judgement based on initial penetration levels of similar products. Using the discrete formulation of the Bass model, the coefficient of external influence, $a$, is equal to the first year penetration levels ($n(1) = aM$, or $n(1)/M = a$, for $N(1)=0$). Purchase intention questionnaires have been found effective in estimating pre-launch penetration levels for the first year of introduction (see Juster 1966; Morrison 1979; Thomas 1985b; Infosino 1986; and Morwitz and Schmittlein 1992). Alternative methods include relying on test markets which are extrapolated to the population, or first-year penetration levels in foreign markets where the innovation has already undergone the initial diffusion process (Gatignon, et al. 1989; Sultan, Farley and Lehmann 1990). Finally, lacking other information, one can simply extrapolate from previous diffusion studies. In a meta-analysis of 15 diffusion studies, or 161 applications of the Bass model, Sultan, Farley and Lehmann (1990) report an average value of .03 for the coefficient of innovation, $a$; Parker (1993a) finds, in an econometric study of 19 consumer electronics categories, that $a$ averages .01; in no case was $a$ estimated to be greater than .10. Caution should be used in relying on these estimates since they typically
do not control for the bias introduced by the significant lag in the reporting of penetration data after the products have been launched; the actual first year penetration levels are likely to be lower than those indicated in previously published empirical studies (Parker 1992, 1993a).

Assuming an initial estimate of the first year and long run penetration levels, the speed of the diffusion process between these two levels and hence the timing of the adoption curve peak must be estimated. Based on the original Bass formulation (e.g. excluding the NUI parameter), a number of estimation procedures have been proposed. Lawrence and Lawton (1981) first suggested that managers make informed guesses of the sum of internal and external influence coefficients (e.g., \(a + b\), in equation 4). While it is not clear that managers (or even academics) can make reasonable guesses, the authors find that this value typically varies between .3 and .7 and suggest that managers rely on estimates from similar products. They recommend, for example, .66 for industrial innovations and .5 for consumer products. Based on this sum, they show how the aggregate diffusion curve can be estimated algebraically given external estimates of the penetration ceiling and the first year penetration level. A similar algebraic approach is suggested by Mahajan and Sharma (1986) who derive the diffusion curve based on informed guesses or delphi estimates of the penetration ceiling, the timing of the adoption curve peak, and the amplitude of that peak. A more structured approach is proposed by Thomas (1985a) based on the systematic comparison of similar products to the innovation being forecasted. Thomas suggests estimating the Bass model parameters using a weighted average of parameters from similar products with diffusion history. Weights reflect the similarity of the new product to analogical products based on 25 criteria across 6 issue areas: including environmental situation, market structure, buyer behavior, marketing strategy, and characteristics of the innovation. Yamada and Leung (1992), using multidimensional scaling of consumer perceptions, weight products based on the reciprocals of the interpoint distances between the innovation and historical products; see also Rao and Yamada 1988. While such approaches have been documented to be used in practice (Mahajan, Muller and Bass 1990), a number of reservations have been raised. First, the algebraic approaches are tantamount to asking managers to simply draw the diffusion curve since managers are asked to guess the key outputs that the models are designed to.
forecast, especially the first purchase peak (Bass 1986). Since delphi or similar techniques can extract managerial guesses for various points along the diffusion curve, it is not clear how much the diffusion models themselves add value to the forecasting process. At a minimum, however, diffusion models have the advantage over naïve alternatives in structuring the problem in a systematic manner. The second concern is that defining "analogical products" may prove difficult or arbitrary, especially for discontinuous innovations. For example, Thomas finds only 3 comparable products for the innovation forecasted as an illustration. The low number of comparable products makes determining the statistical significance of causal assumptions difficult (e.g. determining if certain criteria have resulted in variations in coefficient values). The third major concern is that the models estimated using these procedures have focused on the original Bass specification, proposed in 1969, which has since been shown to not be the best model in characterizing diffusion processes, especially those with asymmetric diffusion patterns, or including the NUI parameter (Easingwood, et al. 1983, Rao 1985, and Parker 1993a).

To overcome this last limitation, Easingwood (1987, 1989) proposes an alternative methodology which incorporates the non-uniform influence parameter, $d$. The procedure involves estimating the two contagion parameters, $b$ and $d$, and clustering curves into classes based on the resulting diffusion curve shapes. In an illustration of the technique, Easingwood and Lunn (1992) derive seven classes of adoption curves, such as rapid take-off and saturation, called "rapid penetration", versus very slow take-off with a high spiked peak, called "late rush". Various criteria, including the launch environment, the target market, company barriers, the innovations' characteristics, and perceived risk are then used to understand the factors influencing why products have particular patterns, and, therefore, values of $b$ and $d$. Based on observed patterns, the new innovation is forecasted based on its characteristics, the market environment, and other relevant criteria. The authors argue that this approach has advantages over the previous approaches since combinations of $b$ and $d$ provide an infinite number of diffusion patterns which are observed in reality. Recently, Balasubramanian and Ghosh (1992) have questioned the reliability of this approach as misclassification (or multiple classification) is likely due to the sensitivity of parameter estimates of $b$ and $d$ to the numbers of observations used from the historical series. They
suggest relying on a joint confidence region for \( b \) and \( d \), rather than point estimates, to classify products. Jones (1992) notes that the procedure is also weakened by the variations in estimates of \( b \) and \( d \), given minor changes in the number of observations used, and, consequently, the inability to correctly/reliably classify products. Jones (1992, p. 353, p. 354) finds that the "parameter estimates are so clouded with variation that Easingwood's classification scheme is rendered inapplicable", and that the "successful implementation of Easingwood's scheme will have to await a better method for estimating parameters of the NUI model."

Another method which incorporates the NUI parameter is proposed in Parker (1993a) which studies the long run diffusion processes of nineteen consumer durable products. Most studies assume an independence across the four diffusion parameters in equation 10 (\( a, b, c \) and \( d \)). If one assumes, however, that realistic pre-launch estimates of \( a \) and \( c \) can be obtained, values of \( b \) and \( d \) can be calculated using the following regression equations which explain 72 percent and 69 percent of the cross-sectional variance in the two parameters respectively (significance levels in parentheses):

\[
b = .6408 + 6.779 a - .4268 c \\
(13) \quad (.0001) (.0002) (.0025)
\]

\[
d = -1.4370 + 5.7119 b \\
(14) \quad (.0104) (.0001)
\]

The equations imply that the higher the long run penetration (e.g. adoption across more segments of society) the lower \( b \), or the initial contagion effect; the higher the initial penetration, the higher the initial diffusion effect. The higher the overall initial contagion effect, the more the diffusion curve is skewed to the left (faster take-off). Interestingly, the initial penetration level, or \( a \), is uncorrelated with the long run absolute penetration ceiling, or \( c \). While these equations may be used to generate priors on \( b \) and \( d \), caution should be used given that the study was limited to nineteen consumer electronics products. A combination of this, and other methods, to obtain priors is likely to provide acceptable priors (Lawrence and Geurts 1984).
In conclusion, pre-launch estimates of two of the basic diffusion model parameters appear plausible due to their direct interpretability: the initial penetration level, $a$, and the ultimate adoption ceiling, $c$. Pre-launch estimates of the remaining parameters ($b$ and $d$) which regulate diffusion speed are less evident given our lack of basic understanding of what the parameters directly reflect. With our current understanding, it is very difficult to explain or attribute the difference between an estimate of $b=0.25$ versus $b=0.35$ in a general sense, even though these two estimates generate radically different diffusion patterns. No study has shown, for example, that categories with high levels of $b$ also historically exhibited high levels of certain types of "word-of-mouth" or other interpersonal communications activities suggested in theory. Even if such relationships were established, it is nevertheless difficult to forecast the likely level of such communication for a new product prior to launch. Promising research avenues will likely involve extensive empirical analyses which systematically determine causal factors (covariates) which can be easily or objectively forecasted prior to launch, and which are reasonable predictors of the relevant diffusion parameters. Such approaches have been used to facilitate diffusion forecasting in a decision support system (called PENETRATOR) which first asks managers to review a number of historical diffusion patterns and answer numerous questions concerning the product concept (Parker 1993b). Based on cross-sectional econometric models applied to over a hundred categories, managers are then provided estimates of long run price and ultimate penetration levels. Based on these, or managerially motivated estimates, and certain assumptions on product breakdown and replacement, diffusion forecasts are generated. Lacking this or similar procedures, highly customized product/industry specific pre-launch modelling approaches may be required. These would rely on extensive industry survey data and managerial judgements; see, for example, Urban, Hauser and Roberts (1990) who describe such a procedure that provides pre-launch forecasts of new automobile sales.

4.2 Post-Launch Forecasting when Few Data are Available

In their review, Mahajan, Muller and Bass (1990) find that "parameter estimation for diffusion models is primarily of historical interest; by the time sufficient observations have been developed for reliable estimation, it is too late to use the estimates for forecasting purposes."
Waiting for enough observations to fit the correct model may render the forecasting exercise a moot issue. This conclusion may be too harsh a generality given that many categories historically have reached their first purchase adoption peaks several decades after product launch due to population growth and slow penetration rates: 20 years for blenders, 22 years for clothes dryers, 23 years for dishwashers, 25 years for disposers, 26 years for freezers, and 20 years for ironers, for example, as shown in Exhibit 2. The total sales peak often occurs, if ever, decades after the first purchase peak due to population growth, and replacement sales. Furthermore, many industry forecasters have demonstrated the importance of post-launch or mature-market diffusion forecasts as market growth uncertainty may remain high even after several post-launch years (see, for example, de Jong, 1980). On the other hand, for many categories the first purchase peak is skewed far to the left (e.g. calculators, black and white television). Heeler and Hustad (1980) and Srinivasan and Mason (1986) suggest that the Bass model is best estimated with a minimum of 10 years' input data for forecasting and market potential estimation. This conclusion should be qualified since (1) these studies were not performed on the more flexible diffusion models (e.g. equation 10), (2) the market potential penetration ceilings were not externally estimated, and (3) parameters were not constrained to plausible ranges (e.g. the first year penetration level equalling the coefficient of external influence). While such constraints may allow for improved estimates with fewer observations, the literature provides insufficient evidence on this subject.

The managerial task during the early years (or months) after launch is to derive as much information from the historical data as possible to update pre-launch parameter estimates. Irrespective of the method used to obtain pre-launch estimates of the diffusion parameters, these priors should be systematically updated as data become available. This is especially critical given that the various diffusion models proposed generate very different forecasts based on relatively minor changes in the parameters estimates; such is the case for most differential equations exhibiting chaotic behavior (see Exhibit 3 and Modis and Debecker 1992). Once data become available (e.g. the first year penetration level), pre-launch estimates of diffusion parameters can be updated using a number of adaptive Bayesian, feedback filter, or parameter updating procedures (see, for example, Bretschneider and Bozeman 1986;
Bretschneider and Mahajan 1980; Lenk and Rao 1989; Sultan, Farley and Lehmann 1990; Vanhonacker 1990; Vanhonacker, Lehmann and Sultan 1990; Vanhonacker and Price 1992; and Wiorkowski 1988). As more observations are used, estimates quickly converge, despite the priors being relatively inaccurate (Lenk and Rao, 1989, pp. 50-51). The following steps illustrate such approaches: (1) obtain "best guess" pre-launch parameter estimates as described above, (2) when the first year penetration level become known, substitute this for \( a \), the coefficient of external influence; based on this new \( a \), update \( b \) and \( d \) using equation 13, or similar relationships based on analogies; keep the long-run penetration ceiling, \( c \), fixed, (3) when the second, third and future year penetration levels become known, fix \( a \) and \( c \) to previous levels (or update \( c \) if new information on product usage becomes available) and update \( b \) and \( d \) using the adaptive Bayesian and feedback procedures discussed above. The ceiling parameter, \( c \), should never be estimated using the early penetration data unless these strongly reveal (with external information) fundamentally different market structures. Likewise, once the first year penetration level is known, \( a \) should be fixed to this level in all subsequent models as this is the most efficient (accurate) use of this information across all of the parameters. In practice, when \( a \) is freely estimated on categories with many decades of data it is often statistically insignificant or has a negative sign (further indicating that the parameter be externally constrained to realistic values; e.g. the first year penetration level). This is especially important since the early years of product diffusion often exhibit chaotic patterns (Modis and Debecker 1992). The usefulness of these earliest data points is very questionable, therefore, to estimate parameters which are responsible for long-run trends.

As more data become available attention should turn to the type of data collected and applied in the updating process. Some confusion over data requirements may be generated by the titles of many studies positioning diffusion models as being simultaneously designed for both "long-run" forecasting (as opposed to time series approaches), and new product "sales" forecasting. For example, the Bass model was proposed as a long-run sales forecasting tool applicable to consumer durable products, as households are likely to adopt only one unit each (adoption data equal sales data). As illustrated for color television by Bayus, Hong and Labe (1989), "sales" can be a far thing from "adoptions". Institutional sales, replacement sales, and multiple unit adoptions by each household can make a substantial difference in sales for many
categories, including consumer durables. This distinction is especially important in the "long run"; for minor innovations or durable products with short life cycles, forecasting need not consider long-run diffusion dynamics. If one is willing to make the assumption that in early years sales equals first adoptions, and forecast the sales peak using an adoption model, the first peak predicted may be one of several to occur in the series, or be dramatically different from the actual sales peak. In a study of repeat purchases of consumer durables, for example, Parker (1991) finds that the Bass formulation unadjusted for replacement sales can be misspecified to track sales for as few as 10 observations, and is dramatically misspecified at 15 observations for most categories. Furthermore, the timing of the "adoption" peak can be many years, even decades, from the actual "sales" peak (Parker 1992). The risk forecasters run in using these models to predict the sales peak may be generated by the mismatch of the model to the sales process.

4.2 Estimation Procedures When Many Data are Available

Provided that the correct model is specified and adequate data are available, a variety of estimation procedures have been proposed in the marketing literature to improve the fit and forecasting ability of the various aggregate diffusion models (Mahajan, Mason, and Srinivasan 1986). These methodologies include ordinary least squares, maximum likelihood estimation, nonlinear least squares, stochastic estimation, Bayesian estimation and other adaptive/feedback estimation procedures.\(^3\) The choice of estimation procedure will largely be driven by the number of observations available, and the functional forms assumed. We briefly review these options by first considering the original Bass (1969) model in equation 4 with a linear error term \(e_i - N(0,\sigma^2)\). When first proposed, Bass (1969) noted that equation 3 can be expanded to make first adoptions a quadratic function of cumulative previous adoptions:

\[
n(t) = aM + (b-a)N(t) - (b/M)N(t)^2
\]  

\[(15)\]

\(^3\) Any number of estimation procedures can be considered, including computer intensive techniques (see, for example, Young and Ord 1985 who consider discounted least squares). Only those explicitly evaluated on diffusion models in the marketing literature are reported.
He suggested (for fitting purposes) that one estimate \( a, b, \) and \( M \) using ordinary least squares (OLS) for the following discrete time model:

\[
nt = \beta_0 + \beta_1 N_{t-1} + \beta_2 N_{t-1}^2 + e_i
\]  

(16)

where \( \beta_0, \beta_1, \) and \( \beta_2 \) are estimated constants and \( e_i \) is the error term assumed \( N(0, \sigma^2) \). Setting \( \beta_0 = aM, \beta_1 = (b-a) \) and \( \beta_2 = (b/M) \) yields three equations with three unknowns which can be solved to obtain estimates of \( a, b, \) and \( M \). When \( M \) is estimated externally the model in equation 16 loses its intercept and \( a \) and \( b \) can be estimated directly. Bass also suggests an algebraic estimation procedure which uses the first three observations in the diffusion series to derive estimates of \( a, b, \) and \( M \); again, this procedure may be hazardous given the chaotic nature of early adoption data (Modis and Debecker 1992).

One shortcoming of the OLS estimation procedure is potentially high levels of multicollinearity; estimates of \( \beta_1, \) and \( \beta_2 \) are unstable. Furthermore, standard errors for the original Bass model coefficients are not obtainable. In noting these shortcomings and the time interval bias introduced when using discrete (usually annual) adoption data, Schmittlein and Mahajan (1982) propose using maximum likelihood estimation. Srinivasan and Mason (1986) propose that nonlinear least squares (NLS) be used to estimate Bass model parameters. Treating equation 4 as a first-order differential equation \( (dN(t)/dt = n(t)) \) and integrating to solve for \( N(t) \) as an explicit function of time, \( t \), one obtains:

\[
N(t) = \frac{1 - e^{(a+b)t}}{1 + (b/a) e^{(a+b)t}} + e_i
\]  

(17)

which can be calibrated using nonlinear least squares while considering both sampling and non-sampling errors (Srinivasan and Mason, 1988; see also Jain and Rao 1990 for specifications incorporating price). Similar formulations exist for a number of the diffusion models proposed in the literature (see Mahajan, Muller and Bass 1990). A desirable feature of this approach is that \( N(t) \) is modelled as an explicit function of time. Parameters can be estimated with their standard errors, and multicollinearity is reduced. Furthermore, this nonlinear estimation approach is found to provide roughly comparable predictive ability to
maximum likelihood estimation. NLS approaches should be applied, therefore, on closed form solutions of aggregate models expressing N(t) as an explicit function of time, if such forms exist, when launch time can be adequately estimated from historical records. In the absence of closed form solutions, NLS can be directly applied to diffusion models. There are, however, two practical limitations to this technique. First, for models which do not have closed form solutions to dN(t)/dt, parameter estimates are not obtainable. This limitation exists for all models including the non-uniform influence (NUI) parameter (d); a parameter which appears important in characterizing long-run diffusion processes. In such cases, nonlinear least squares can be directly applied to the aggregate diffusion model which is not formulated explicitly as a function of time (e.g. equation 10); parameter standard errors are obtainable in such cases, but N(t) is not modelled as an explicit function of time. The second limitation of using the closed form solution to N(t) is that the origin for time, t, may not be apparent for the innovation studied. In theory, t should begin when the product is "launched". In many cases, this may be an arbitrary date, and will necessarily affect forecasts of the peak timing. For example, color televisions were first sold in the United States in 1955, yet most diffusion modellers consider 1960, 1961, or 1962 as being the launch year (e.g. after standardization existed). Typically diffusion modellers choose the launch date as equalling the year of data availability. This approach can systematically overstate the launch date by several years, or even decades (Parker 1992, p. 361), as trade associations or the trade press which report diffusion data do so only after a product category is perceived to be long-lasting. Additional limitations of the NLS procedure include non-converging estimates, and local minima; these can typically be handled using wide area initialization of parameters. In their review, Mahajan, Muller and Bass (1990, p.9) conclude that various estimation methods "often yield estimates that do not differ greatly" especially when a large number of historical data are available. While nonlinear least squares appears to offer the best general approach to aggregate diffusion model estimation, Dalal and Weerahandi (1993) propose a maximum likelihood estimation procedure with beta binomial approximation and find their approach superior to NLS and weighted NLS, especially in terms of forecasting the long-run market potential. Their study is limited, however, to a single category, which is applied to the original Bass formulation.
More important than the estimation procedure used are the constraints one places on the parameter values based on external theoretical considerations. Even when a large number of observations are available, researchers should not hesitate imposing constraints on parameters as strong theories generate substantial, though non-statistical, degrees of freedom. In particular, the modeller may find it useful to impose the first year penetration level, if this is available, as an estimate of the external influence coefficient, \( a \); the penetration ceiling, \( c \), should also be constrained to external estimates or at a minimum be bounded to be no greater than 1.0. There are clear advantages in estimating \( c \) as a parameter (as opposed to estimating a fixed \( M \), the market potential in absolute units). First, \( c \) has more direct interpretation as a unit of measure (a percentage) and is bounded to a known range. When \( M \) is estimated internally and directly, as in equations 4 or 15, it can typically indicate large absolute numbers (e.g. in units reaching the thousands or millions) which are difficult to verify based on managerial perception; a percent of a target market (or share gained from existing substitutes) can be more easily gauged for face validity (e.g. percentages greater than 1.0 indicate specification, data collection, or market definition errors, whereas a large estimate of \( M \) may not signal these problems). Second, since forecasts are long-run in nature, \( M \) must be dynamic (or allowed to be so), so externally measuring \( M(t) \) as a variable is required. Using \( M(t) \) as a variable forces the modeller to understand and objectively identify the market (social system) within which the innovation is undergoing a diffusion process, and adapt data collection and modelling procedures accordingly. In this sense, modellers who do not impose such constraints and allow the model to "do all of the talking", may not have thought enough about how the diffusion process and the model should be matched. Finally, the NUI parameter, \( d \), can be constrained to be greater than -1.0, while the coefficient of internal influence, \( b \), can be constrained to be positive (if the NUI parameter \( d \) is excluded, then \( b \) should vary between 0 and 1.0; if NUI is included, \( b \) can actually take on values greater than 1.0 since it acts as a multiplicative intercept to a term which is less than one, \( b(N(t)/M(t)^d) \)). While the usual statistical caveats apply when using constrained estimation, without such constraints, diffusion models frequently generate implausible estimates, poor long-run forecasts, or incorrectly indicate premature market saturation due to short term business cycle effects, among others. Since estimates mostly reported in the literature are unconstrained, further empirical work on the effects or potential benefits of such constraints is warranted.
5. Future Research Directions

Simon's (1992) conclusion that diffusion models are risky and misleading as forecasting tools illustrates the current state of affairs for many new product modellers. This review has illustrated that the use of these models as a generic tool may be asking too much. Like many management science techniques, diffusion forecasting requires a thorough understanding of both the diffusion process and the market situation before model specification and estimation is undertaken. In their general review, Mahajan, Muller and Bass (1990, p. 21) call for greater empirical work to help "identify conditions under which diffusion models work or do not work" in forecasting situations. Despite the managerial relevance and substantial academic interest in this area over the past 30 years, the proliferation of models and methodologies leaves forecasters with few "rules-of-thumb" or empirical generalizations to work with, which may lead to frustration in applying these models in practice. While there has recently been responses to this call in terms of understanding appropriate parsimonious specifications (Rao 1985, Parker 1993a), we are only beginning to understand how these models should be optimally implemented, especially prior to product launch. To conclude this paper, three future research directions are elaborated upon that should provide more basic or risk-reducing insight on these models: (1) more rigorous and systematic evaluations of diffusion models as applied forecasting tools, (2) research on the basic validity of the diffusion parameters themselves and studies on the factors determining their levels, and (3) extensions of this research to the study of forecasting international or geographic diffusion processes.

With respect to the first research area, diffusion studies in marketing typically de-emphasize forecasting issues; most diffusion studies fail to systematically compare proposed diffusion models to reasonable or more simplistic alternatives. Mahajan, Sharma and Bettis (1988), for example, have found that alternative time-series approaches can also generate s-shaped curves and may be more applicable than diffusion models. The exclusion of non-diffusion approaches in empirical diffusion studies opens a number of important applied and academic research issues:
- other than the issue of forecasting the first purchase peak, there is a general lack of knowledge on the comparative advantage of diffusion models over time series approaches, and whether a combination of approaches is desirable; a general analysis evaluating the impact of data availability, the forecast horizon, and the number of parameters included in diffusion models is warranted;

- with the exception of the original Bass (1969) model, there is no general understanding of the number of data required to obtain stable parameter estimates, especially for the NUI parameter, and whether this number typically occurs prior to or after the peak in adoption (or sales, in the case of repeat purchase models); recursive coefficient estimation procedures should prove useful in this regard;

- it is generally accepted that diffusion parameters show non-stationarities as more observations are used in estimation. Little is generalizable concerning the extent of this phenomenon, and whether there are systematic patterns of instability (e.g. as revealed in recursive regression) which can be accounted for when forecasts are developed using few historical data; time varying parameter approaches have been suggested to capture such changes when sufficient data are available (Mahajan, Bretschneider and Bradford 1980).

Furthermore, most diffusion modellers de-emphasize difficulties in adequately measuring the impact of various foreseen events on the diffusion process. These might include the lifting of supply restrictions (Simon and Sebastian 1987, or Jain, Mahajan and Muller 1991), or competitive entry (Parker and Gatignon 1992, 1994). Such studies will assist managers in distinguishing between an early dip in first purchases due to a business cycle, versus market saturation. As the number of forces are potentially infinite, decision calculus, or delhi procedures appear to be the best alternatives available, especially in pre-launch forecasting. The literature tends to suggest that a four parameter model (or three if parsimony is desired) can guarantee an adequate fit/forecast to diffusion data in practice (as is the case for widely accepted s-shaped sales response models used in marketing decision calculus applications; see CALLPLAN, Lodish 1971). While the 4-parameter diffusion model in equation 10 provides a rather intuitive and rigorous framework within which forecasters can structure the new
product forecasting problem, future researchers should invest greater efforts in generalizing knowledge on the forecasting behavior of these models and in evaluating conditions under which these models (under various estimation scenarios) will typically outperform naïve alternatives likely to be used in their place (Kolb and Steckler 1993). This knowledge can then be the basis upon which managers can develop reasonable forecasts and explore the sensitivities of underlying assumptions being made. Most diffusion studies fall short of providing such insights, and this is often the result of basic data handling problems:

- despite a wealth of data and the context-free environment for which these models have been developed, proposed models generally are presented with a limited number of products (from 3 to 6) representing a convenience sample within a single industry; the basis upon which categories are selected is generally not stated (other than citing previous use in the literature); despite the existence of adoption, first purchase, or trial data, proposed models often use early sales data without controlling for repeat purchases; likewise, the use of production data, which do not control for exports, may lead to poor fits of the Bass model (Heeler and Hustad 1980);

- without recognition/acknowledgement, time series appear to be truncated on an ad hoc basis; the left hand tail is truncated, sometimes due to data unavailability, which masks the actual launch date, the years to the adoption or sales peak and the skew of the adoption curve; the right hand tail (pre-and post peak) is often truncated which can modify the perceived or predicted/fit first peak which is often not the unique or highest sales peak, producing exaggerated fit statistics; this "right-hand" truncation may not be wise from a managerial point of view in that first purchases can represent from 20 to over 40 percent of sales even several decades after launch for many consumer durables (Parker 1992);

- given the dependence of diffusion processes on external variables (marketing mix, competitive dynamics), studies fail to report the forecasting issues (and the possible compounding of errors) associated with including these independent variables.

On this last point, Bass and Krishnan (1992), in noting that parsimonious aggregate models fit diffusion data well, demonstrate that the exclusion of various market mix and exogenous
factors in aggregate diffusion models may not be a serious problem when these factors do not affect the long-run penetration ceiling and enter the equation in a general multiplicative manner.

Returning to Rogers (1983), marketing diffusion models have concentrated on two of the four basic elements of diffusion processes: communication channels (interpersonal influences) and time. It is clear that diffusion modelling has significantly progressed in this area and that less attention should be placed on developing more complex functional forms which capture the basic s-shape function. The two remaining elements, however, have been under-researched: the innovation and the social system. An important avenue for future research will be the search for empirical generalizations across products and social systems which can guide forecasters in using diffusion models at the earliest phases of product planning (Bass 1993). These studies encompass the two remaining areas of proposed research. With respect to the innovations, Roberston (1969, 1971) and Gatignon and Robertson (1985) have suggested that post-launch diffusion patterns are in part driven by the types of products, and their perceived attributes. Cross-sectional studies should be made to further understand parameter variations across products. These would answer basic questions, for a given industry: (1) what features or benefits of a product typically generate certain diffusion patterns?; (2) if products have already reached a certain level of acceptance, what will be their likely acceptance in the future?; and (3) are parameters reflecting the diffusion process interdependent? This research should cross-sectionally incorporate physical, perceptual and social dimensions of products. The field of diffusion modelling will be further served by the publication of more widespread empirical studies across a much broader class of products purchased in a variety of marketing contexts. Armed with more general knowledge, pre-launch forecasts can be justified based on tested theories of diffusion theory (Gatignon and Robertson 1985), as opposed to ad hoc manipulations of data from existing categories. Work in this area has been hampered by the lack of large or standardized data sets which can be competitively analyzed by various authors, as is the tradition in other forecasting arenas. The advent of optical scanner data may prove useful in this regard. Even though they have been traditionally applied to consumer durables, there is absolutely no reason to avoid using diffusion models to analyze the trial and repeat purchase processes of consumer packaged goods, as many studies have documented
the importance of interpersonal influence for such categories (see the review in Dawar, Parker and Price 1993). Since managers are equally interested in their individual brands’ diffusion rates, analyses of brand-level scanner data should be encouraged, especially if these can be used on a mass scale in order to generate basic understanding.

Finally, with respect to social systems, in their recent review of the international marketing literature, Douglas and Craig (1992) found limited international diffusion research. As it stands, diffusion parameters generally lack external validity in terms of us understanding what these actually measure (e.g. actual word-of-mouth communications, versus supply constraints; Simon and Sebastian 1987). While some research has recently occurred in this area (e.g. Bundgaard-Nielsen 1975; Gatignon et al. 1989; Helsen et al. 1993; Takada and Jain 1991), we are at the earliest stages of general understanding. For example, we may observe that the Japanese have different diffusion parameters than other social systems, yet we are not sure whether this is due to different perceptions of products, different uses of products, different supply constraints, or different aversions to risk. Since most products are not launched simultaneously in multiple countries (or geographic regions), large-scale international studies using data from multiple categories should allow forecasters to anticipate the diffusion pattern of a product globally, or by country individually, given a historic pattern in the home market. Gatignon et al., 1989, for example, discuss how estimates of diffusion parameters obtained on one country can be used to estimate adoption rates in countries which have no historical data. The lessons learned from these and future studies will greatly improve the usefulness in reducing the potential risks in using diffusion models as applied forecasting tools.
REFERENCES


Mahajan, V. and E. Muller (1979), "Innovation Diffusion and New Product Growth Models in Marketing," *Journal of Marketing*, 43 (Fall), 55-68.


Exhibit 1. Category Penetration Level versus Category Age (in years)
Exhibit 2
Annual First Purchase Patterns Over Time for Consumer Durable Goods
(solid lines are actual observations; dashed lines are fitted using Equation 10)
Exhibit 3
Sensitivity of Diffusion Patterns to Changes in Parameters

1. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.25$
2. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.25$
3. $a_i = .0001$, $b_i = .4$, $c_i = .4$, $d_i = -.25$
4. $a_i = .001$, $b_i = .4$, $c_i = .4$, $d_i = -.25$

$a_i$ varies

1. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.25$
2. $a_i = .00001$, $b_i = .2$, $c_i = .4$, $d_i = -.25$
3. $a_i = .00001$, $b_i = .6$, $c_i = .4$, $d_i = -.25$
4. $a_i = .00001$, $b_i = .99$, $c_i = .4$, $d_i = -.25$

$b_i$ varies

1. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.25$
2. $a_i = .00001$, $b_i = .4$, $c_i = .1$, $d_i = -.25$
3. $a_i = .00001$, $b_i = .4$, $c_i = .7$, $d_i = -.25$
4. $a_i = .00001$, $b_i = .4$, $c_i = .99$, $d_i = -.25$

$c_i$ varies

1. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.25$
2. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.10$
3. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.50$
4. $a_i = .00001$, $b_i = .4$, $c_i = .4$, $d_i = -.75$

$d_i$ varies