ELSEVIER

Intern. J. of Research in Marketing 23 (2006) 93 – 106

International Journal of Research in Marketing

www.elsevier.com/locate/ijresmar

# Virtual Bass Model and the left-hand data-truncation bias in diffusion of innovation studies

Zhengrui Jiang [a], Frank M. Bass [b,*], Portia Isaacson Bass [c]

[a] College of Business, University of North Alabama, UNA Box 5199, Florence, AL 35632, USA
[b] School of Management, The University of Texas at Dallas, P.O. Box 830688, Richardson, TX 75083-0688, USA
[c] Bass Economics, Inc., 4 Glenmoor, Frisco, TX 75034, USA

**Abstract**

One of the most important applications of the Bass model is "guessing by analogy" — the parameter estimates of the Bass model for analogous products can be used to predict the diffusion pattern of a new product. However, estimates based on left-hand truncated data will be biased unless care is taken to adjust for the bias. We demonstrate the prevalence of left truncation in historical sales data and present a method for dealing with the bias by the *Virtual Bass Model*. We use the model to develop a database of parameter estimates that may be used in "guessing by analogy."


*Keywords:* Bass model; Diffusion of innovations; New product forecasting; Left-hand truncation bias; Guessing by analogy

## 1. Introduction

Several studies of estimation issues associated with the Bass model (Bass, 1969) have taken note of difficulties in estimation associated with data that are right truncated. Putsis and Srinivasan (2000) in a review paper of estimation methodology have indicated that stable and robust parameter estimates can be obtained only if the data include the peak of the adoption curve. Van den Bulte and Lilien (1997) in a study utilizing both empirical and simulated data claim that estimates of the parameters of the Bass model are biased and that the amount of bias is affected by the extent to which the data are right censored. Bemmaor and Lee (2002) have studied the effects of "extra Bass" skew in the data using both simulated and real adoption data sets and conclude that the parameter estimates depend on "extra skew" in the data and the level of right censoring. Although estimation issues concerning right censored data have been examined extensively, left-hand truncation has been mostly ignored, even though the great majority of diffusion studies have used data starting after the true launch time of the innovation. Many researchers are aware that data series often suffer from left truncation, but only limited attempts have been made to address the issue. As a result, almost all published estimates of the Bass model suffer from left-hand truncation bias. This is very problematic for people seeking to make pre-launch forecasts using prior results pertaining to analogous products. The present paper develops a procedure to correct for left-hand truncation bias in published estimates of the Bass model parameters. We will demonstrate here the use of the Virtual Bass Model in correcting for left-hand truncation bias and show that this method is superior to existing alternative methods.

From the outset of the publication of the Bass model paper there was the notion that the model provided a framework for "guessing without data." In that article (Bass, 1969) one finds the following quotation: "Long-range forecasting of new product sales is a guessing game, at best. Some things, however, may be easier to guess than others. The theoretical framework presented here provides a rationale for long-range forecasting." Soon after the publication of the Bass model paper in 1969 guessing algorithms built around the framework of the Bass model were developed. One of these algorithms was developed and used extensively at Eastman Kodak and was also used by several other companies including RCA, IBM, Sears, and AT and T. This procedure is described by Lawrence (1981) and by Lawrence and Lawton (1981). Putsis and Srinivasan (2000) have reviewed various algorithms that have been employed in guessing parameter values when little or no data are available.

* Corresponding author. Tel.: +1 972 883 2744; fax: +1 972 883 2799.
*E-mail address:* mzjb@utdallas.edu (F.M. Bass).

0167-8116/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijresmar.2006.01.008

Over time the idea of "guessing by analogy" has emerged as a basis for forecasting sales of new products and technologies prior to launch. Extensive databases of Bass model parameter estimates of previously introduced products have been developed. In the guessing-by-analogy method the $p$ and $q$ parameter estimates for a (previously introduced) analogous product are used to forecast the diffusion pattern for the new product. Lilien, Rao, and Kalish (1981) used this method to forecast sales of a new prescription drug and Bayus (1993) used a segmentation scheme to group "similar" products and generate forecasts for high-definition television (HDTV). An article by Bass, Gordon, Ferguson, and Githens (2001) provides a case history of the successful use of the guessing-by-analogy method in forecasting the diffusion of satellite television prior to launch.

Although the guessing-by-analogy method has proven to be useful, there is a potentially serious problem with this method in some cases. The problem arises because in many instances the available historical data start several years after product introduction. Estimates based on left-hand truncated data series will be biased unless care is taken to adjust for the bias. For products where there is a long left tail of the plot of sales over time the bias will be substantial. Other authors have taken note of the importance of the left-hand truncation bias problem. These include Dekimpe, Parker, and Sarvary (1998), Dekimpe, Parker, and Sarvary (2000), Kohli, Lehmann, and Pae (1999), Bayus (1992), Van den Bulte (2000), and Lilien, Rangaswamy, and Van den Bulte (2000). Meta-analysis studies by Sultan, Farley, and Lehmann (1990) and Vanhonacker, Lehmann, and Sultan (1990) have compiled Bass model parameter estimates for several product categories. In addition, Bass model parameter estimates for several new products have been published in papers by Kohli et al. (1999) and Bayus (1992). Unfortunately, for many of the product categories included in these studies sales data are not available from the time of product introduction; therefore, in many cases the parameter estimates for these products suffer from left-hand truncation bias. Lilien et al. (2000), aware of the left-truncation issue, used penetration data in conjunction with the discrete-time version of the Bass model differential equation in an effort to avoid left-truncation bias. We shall later compare their approach with the method that we develop here. We also examine the possibility suggested by a reviewer that the Mahajan and Peterson (1985) solution to the Bass model differential equation can be used to adjust for left truncation. As our final alternative, we consider the possibility of treating the difference between the product introduction time and the start of the data as a parameter to be estimated.

In this paper we will demonstrate the prevalence of the left-hand truncation bias and present a method for correcting the bias by exploiting previously unknown mathematical properties of the Bass model. We call this extension of the Bass model the *Virtual Bass Model.* We show that for any Bass diffusion process with a given start time there is an equivalent process for any other start time. The Virtual Bass Model is of both theoretical and practical significance. In particular, we will show how to obtain the correct parameter estimates when the data are left-truncated. This permits us to develop a correct set of parameter values for guessing by analogy.

In the next sections we develop the theoretical bases for the Virtual Bass Model.

## 2. The Bass model is symmetric about $T^*$

We begin with a proof that if $q>p$, the Bass diffusion curve over the interval between 0 and $2T^*$ is symmetric with respect to $T=T^*$, where $T^*$ is the time of peak sales. We use this result in the theoretical developments to follow.

We use the essential Bass model equations in the proof and thus we repeat their development below. The Bass model is derived from a simple premise that the conditional likelihood of adoption of a randomly chosen consumer at time $T$, given that adoption has not yet occurred, is a linear function of the number of previous adopters. The mathematical form of the assumption is shown in the hazard function of Eq. (1):

$$f(T)/[1-F(T)] = p + (q/m)Y(T), \tag{1}$$

where $m$, $p$, and $q$ are constant parameters representing the total number of potential adopters, the coefficient of innovation, and the coefficient of imitation, respectively; $F(T)$ and $f(T)$ are the cumulative and non-cumulative proportion of adopters at time $T$; $Y(T)$ is the total number of adopters by time $T$. The diffusion rate $S(T)$ at time $T$ can be expressed as Eq. (2) or, equivalently, as Eq. (3) and the total number of initial adoptions by time $T$ is shown in Eq. (4):

$$S(T) = pm + (q-p)Y(T) - (q/m)[Y(T)]^2, \tag{2}$$

$$S(T) = \frac{m(p+q)^2}{p}\frac{e^{-(p+q)T}}{[(q/p)e^{-(p+q)T}+1]^2}, \tag{3}$$

$$Y(T) = \int_0^T S(t)\mathrm{d}t = \frac{m(1-e^{-(p+q)T})}{(q/p)e^{-(p+q)T}+1}. \tag{4}$$

The shape of the Bass diffusion curve varies according to two different conditions. If $q \le p$, the diffusion rate decreases monotonically with time $T$ and if $q>p$, $S(T)$ increases monotonically before peak time, $T^*$, and drops monotonically afterwards. The peak time can be calculated based on the following equation:

$$T^* = [1/(p+q)]\ln(q/p). \tag{5}$$

Another important property of $S(T)$ is stated in Theorem 1 and shown in Fig. 1.

**Theorem 1.** *If $p<q$, the Bass diffusion curve segment over [0, $2T^*$] is symmetric with respect to $T=T^*$.*[1]

**Proof.** To prove that the Bass diffusion curve segment over [0, $2T^*$] is symmetric, we only need to show that for any $0 \le t \le T^*$,

[1] Mahajan and Wind (1986) indicated that the Bass model assumed symmetry and Mahajan, Muller, and Srivastava (1990) noticed that the Bass model was symmetric between 0 and $2T^*$. However, no proof of the symmetry property was offered in either of these papers. We believe that this is the first published formal proof of symmetry of the Bass model.

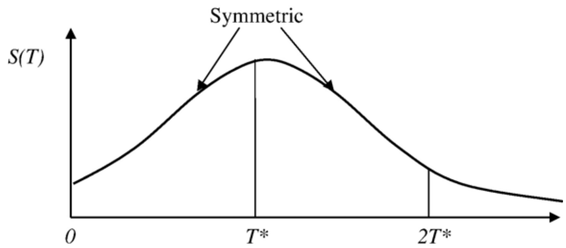


Fig. 1. Symmetry of bass diffusion curve.

$S(T^*-t)=S(T^*+t)$ always holds. From Eq. (3), we have:

$$S(T^*\pm t) = \frac{m(p+q)^2}{p}\frac{e^{-(p+q)(T^*\pm t)}}{[(q/p)e^{-(p+q)(T^*\pm t)}+1]^2} = \frac{m(p+q)^2}{p}\frac{k\cdot e^{\mp(p+q)t}}{[e^{\mp(p+q)t}+1]^2},$$

where

$$k = e^{-(p+q)T^*} = p/q.$$

Noting that for any real $x$ (thus in particular, for $x=(p+q)t$), $\frac{e^{-x}}{[e^{-x}+1]^2}=\frac{e^{-x}\cdot e^{2x}}{[e^{-x}+1]^2\cdot e^{2x}}=\frac{e^{+x}}{[e^{+x}+1]^2}$, which completes the proof. □

## 3. Virtual Bass Model

We show in this section that by a novel way of constructing a Virtual Bass Model (or VBM), the symmetry property for the Bass model holds in $(-\infty, +\infty)$. A Virtual Bass Model is represented by a hypothetical Bass diffusion curve that starts from time negative infinity $(-\infty)$, which can be obtained by extending a known Bass diffusion curve to time $-\infty$. Before this can be done, we first introduce two ways of transforming the Bass model: *forward transformation* and *backward transformation*. We then derive the *Virtual Bass Model Equations (or VBE)*, which makes the computation of transformed parameters simpler without going through the forward and backward transformations. Finally, we show how parameters for a backward transformed diffusion process can be calculated based on VBE.

### 3.1. Forward transformation

An in-progress Bass diffusion process $D_0$ at time $T=\tau$ can be transformed into a new process $D_F$ for the remaining portion of the potential adopters. The transformation, which we term forward transformation, is illustrated in Fig. 2. We denote the three parameters of $D_0$ by $m$, $p$, and $q$ and those of $D_F$ by $m'$, $p'$, and $q'$. Since the start times of the two processes are different, we denote the timing relative to the start time of $D_0$ by $T$ and that for $D_F$ by $T'$, with $T'=(T-\tau)$. In addition, we use $S(T)$ and $S'(T')$ to represent the diffusion rate functions for $D_0$ and $D_F$, and $Y(T)$ and $Y'(T')$ to represent the cumulative number of adoptions for $D_0$ and $D_F$, respectively.

From the forward transformation construction, the following must hold for any $T\geq\tau$ or $T'\geq 0$:

$$S(T)\equiv S'(T') \quad (6)$$

$$Y'(T')\equiv Y(T)-Y(\tau), \quad (7)$$

where $Y(\tau)$ is the number of adopters of $D_0$ at $T=\tau$ and can be obtained based on Eq. (4).

Assuming that the three parameters for $D_0$ are known, we can derive the parameters for $D_F$. By letting $T$ (and thus $T'$) in Eq. (7) go to positive infinity, we get the total market potential for $D_F$:

$$m' = m-Y(\tau). \quad (8)$$

By expanding both sides of Eq. (6) based on Eq. (2), we have:

$$pm + (q-p)Y(T)-(q/m)[Y(T)]^2 = p'm' + (q'-p')Y'(T')-(q'/m')[Y'(T')]^2. \quad (9)$$

After plugging Eq. (7) into Eq. (9) and algebraic rearrangement, we obtain the following:

$$pm + (q-p)Y(T)-(q/m)[Y(T)]^2 = p'm'-(q'-p')Y(\tau)-(q'/m')[Y(\tau)]^2 + [(q'-p') + 2(q'/m')Y(\tau)]Y(T)-(q'/m')[Y(T)]^2. \quad (10)$$

Since Eq. (10) must hold for every $T$, the following three equations must also hold:

$$q/m = q'/m', \quad (11)$$

$$q-p = (q'-p') + 2(q'/m')Y(\tau), \quad (12)$$

$$pm = p'm'-(q'-p')Y(\tau)-(q'/m')[Y(\tau)]^2. \quad (13)$$

From Eqs. (11) and (8), we obtain $q'$:

$$q' = q-(q/m)Y(\tau). \quad (14)$$

Based on Eqs. (11), (12), and (14), we derive the solution for $p'$:

$$p' = p + (q/m)Y(\tau). \quad (15)$$

Substituting Eqs. (8), (14) and (15) into Eq. (13), we find that the equality in Eq. (13) holds. Therefore, Eqs. (8), (14) and (15) combined constitute the unique solution for forward transformation. The solution guarantees $S(T)=S'(T')$ for any $T$. Based on Eqs. (14) and (15), we further conclude that $p'+q'=p+q$ is always true for forward transformation.

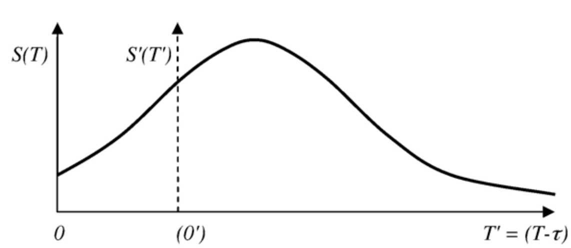


Fig. 2. Diffusion process forward transformation.

Table 1
Parameters in backward transformation

| Backward transformed process ($D_B$) | Original process ($D_0$) |
|---|---|
| $T'$ | $T=T'-\tau$ |
| $Y'(T')$ | $Y(T)=Y'(T')-Y'(\tau)$ |
| $m'$ | $m=m'-Y'(\tau)$ |
| $q'$ | $q=q'-(q'/m')Y'(\tau)$ |
| $p'$ | $p=p'+(q'/m')Y'(\tau)$ |
| $S'(T')=S(T)$ | |

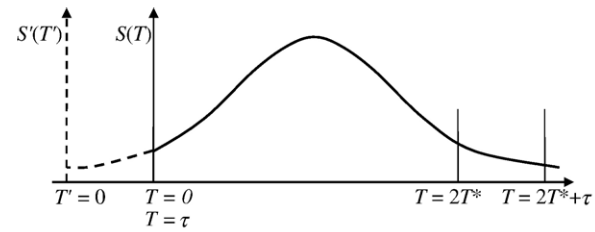


Fig. 3. Direct backward transformation.

For an in-progress Bass diffusion process $D_0$ the bias in the parameters estimated for a process that starts at $\tau$ (using Eqs. (8), (14), and (15)) will be:

$$m'-m = Y(\tau),$$
$$q'-q = -(q/m)Y(\tau),$$
$$p'-p = (q/m)Y(\tau).$$

As a special case of forward transformation, if $p<q$ and $\tau=2T^*$, the parameters for the forward transformed diffusion $D_{Fs}$ can be derived as follows:

From $e^{-(p+q)2T^*}=p^2/q^2$, we have:

$$Y(2T^*) = \frac{m(1-p^2/q^2)}{(q/p)(p^2/q^2)+1} = m(1-p/q).$$

Based on Eqs. (8), (14) and (15), we obtain:

$$\begin{cases} m_{Fs} = m-Y(2T^*) = (p/q)m, \\ q_{Fs} = q-(q/m)Y(2T^*) = p, \\ p_{Fs} = p-(q/m)Y(2T^*) = q. \end{cases} \quad (16)$$

We name $D_{Fs}$ *the forward transformed symmetric diffusion process of* $D_0$ since the starting times for $D_0$ and $D_{Fs}$ are of equal distance to $T^*$.

### 3.2. Backward transformation

The reverse of forward transformation is *backward transformation*. In the forward transformation just discussed, if we assume the known diffusion process is $D_F$, $D_0$ can be considered the backward transformed process of $D_F$. In backward transformation, we refer to the original process again as $D_0$ and the transformed process as $D_B$. We still denote the three parameters of the original process by $m$, $p$, $q$ and those of the transformed process by $m'$, $p'$ and $q'$. Since $D_0$ can be considered the forward transformed process of $D_B$, based on the solution of forward transformation, we express the parameters for $D_0$ in terms of the parameters for $D_B$, as shown in Table 1. After some algebraic rearrangements, we obtain the parameters for the backward transformed process, as shown in Table 2. Since this solution is derived from forward transformation, the equality $S(T)=S'(T')$ still holds for any $T$.

Table 2
Parameters rearranged for backward transformation

| Original diffusion ($D$) | Transformed process ($B$)[a] |
|---|---|
| $T$ | $T'=T+\tau$ |
| $Y(T)$ | $Y'(T')=Y(T)+Y'(\tau)$ |
| $m$ | $m'=m+Y'(\tau)$ |
| $q$ | $q'=q+(q/m)Y'(\tau)$ |
| $p$ | $p'=p-(q/m)Y'(\tau)$ |
| $S(T)=S'(T')$ | |

[a] In case of $p\le q$, $Y'(\tau)=Y(2T^*+\tau)-Y'(2T^*)$.

Unlike in forward transformation, $Y'(\tau)$ in backward transformation is not readily available. The solution, therefore, cannot be directly used unless $p\le q$. As illustrated in Fig. 3, in the case of $p\le q$, based on the symmetric property of the Bass model, we have:

$$Y'(\tau) = Y(2T^*+\tau)-Y(2T^*). \quad (17)$$

We call the solution shown in Fig. 3 for the $p\le q$ case *direct backward transformation.*

In the case of $p>q$, the problem can be solved through *indirect backward transformation*. With indirect backward transformation, we first find the *backward transformed symmetric diffusion process* $D_{Bs}$, and then use $D_{Bs}$ to obtain $D_B$, the requested backward transformed process of $D_0$. Analogous to general backward transformation, backward symmetric transformation can be considered the reverse of forward symmetric transformation discussed in the previous subsection; therefore, $D_0$ is the forward transformed symmetric diffusion process of $D_{Bs}$. The $D_{Bs}$ parameters are derived (see Table 3) by following a similar algebraic transformation (see Tables 1 and 2). In both forward and backward symmetric transformation, the values of $p$ and $q$ for the original process are switched in the transformed symmetric process.

The distance between the start times of $D_{Bs}$ and $D_0$, denoted by $\theta$, can be calculated with:

$$\theta = 2(T_{Bs})^* = [2/(p_{Bs}+q_{Bs})]\ln(q_{Bs}/p_{Bs})$$
$$= [2/(p+q)]\ln(p/q).$$

To obtain the backward transformed process $D_B$ from $D_{Bs}$, we first compare $\tau$ with $\theta$ and then follow the following rule:

Table 3
Parameters rearranged for backward transformation

| Original diffusion ($D_0$) | Transformed diffusion ($D_{Bs}$) |
|---|---|
| $m$ | $m_{Bs}=(p/q)m$ |
| $q$ | $q_{Bs}=p$ |
| $p$ | $p_{Bs}=q$ |

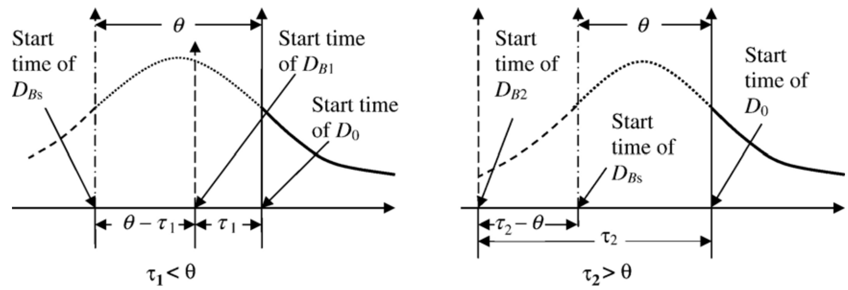


Fig. 4. Indirect backward transformation ($p \leq q$).

(1) If $\tau=\theta$, $D_B$ is the same as $D_{Bs}$.
(2) If $\tau<\theta$, $D_B$ can be obtained by forward transforming $D_{Bs}$ by $(\theta-\tau)$ units of time.
(3) If $\tau>\theta$, $D_B$ can be obtained by backward transforming $D_{Bs}$ by $(\tau-\theta)$ units of time.

Fig. 4 illustrates the indirect transformation with two examples where $D_{B1}$ is obtained by forward transforming $D_{Bs}$ by $(\theta-\tau_1)$ units of time since $\tau_1$ is less than $\theta$, and $D_{B2}$ is obtained by backward transforming $D_{Bs}$ by $(\tau_2-\theta)$ units of time since $\tau_2$ is greater than $\theta$. The indirectly obtained backward transformed processes still satisfy the condition that $S(T)=S'(T')$ for $T' \geq \tau$ or $T \geq 0$, and the symmetric property still holds if $p'<q'$.

It can be verified that regardless of the condition in the original process, the following equation still holds in backward transformation: $(p'+q')=(p+q)$.

### 3.3. The Virtual Bass Model

By backward transforming a Bass diffusion curve to negative infinity, we can imagine that there exists a virtual diffusion curve that spans $(-\infty, +\infty)$, as shown in Fig. 5. We call this a *Virtual Bass Model* (or VBM). Along a VBM curve, once a starting point is fixed, a specific Bass diffusion process is known and its three parameters $m$, $p$, and $q$ are fixed. The VBM Curve is bell-shaped, unimodal, and symmetric in $(-\infty, +\infty)$ with respect to $T=T^*$.

Another important property of the VBM is stated in the following theorem:

**Theorem 2.** *Any segment of a Bass diffusion curve uniquely determines a VBM curve.*

**Proof.** In Appendix A. □

**Corollary 1.** *The solutions for the forward and backward transformations are unique solutions.*

**Proof.** We have shown that a forward or backward solution satisfies $S(T)=S'(T')$. For any curve segment to the right of the start time of the rightmost process, $D_F$ in forward transformation or $D_0$ in backward transformation, the transformed process matches the curve segment perfectly. From Theorem 2, this must be the unique solution. □

From the solutions for forward and backward transformation we obtain parameters for a Bass diffusion process starting from $-\infty$, $+\infty$, or $T^*$.

(I) In order to get the parameters for the diffusion process that starts from $-\infty$, we let $\tau$ go to $\infty$ in the backward transformation. If $p$ is less than $q$, from Table 2, we have:

$$Y'(\infty)=Y(\infty)-Y(2T^*)=m-m(1-p/q)=mp/q,$$
$$\begin{cases} m'=m+Y'(\infty)=m+mp/q=(1+p/q)m, \\ q'=q[m+Y'(\infty)]/m=p+q, \\ p'=p-(q/m)Y'(\infty)=0, \end{cases} \quad \text{in case of } p \leq q. \tag{18}$$

If $q$ is less than $p$ in the known diffusion process, we can get the following parameters indirectly based on Eq. (18) and the parameters for the backward transformed symmetric process $B_s$:

$$\begin{cases} m'=(1+p_{Bs}/q_{Bs})m_{Bs}=(1+p/q)m, \\ q'=p_{Bs}+q_{Bs}=p+q, \\ p'=0, \end{cases} \quad \text{in case of } p>q. \tag{19}$$

The expressions for the three parameters are the same in Eqs. (18) and (19).[2] For convenience, we introduce two more parameters, $M$ and $U$, and let $M \equiv (1+p/q)m$ and $U \equiv (p+q)$. Since the parameter $m$ for any Bass diffusion process is interpreted as the total number of potential adopters and equals the integral of $S(T)$ from time zero to infinity, and the Bass diffusion curve that starts from $-\infty$ is the same as the VBM curve, $M$ can be interpreted as the *total number of virtual adopters* of a VBM and equals the integral of the VBM curve from $-\infty$ to $+\infty$. From Theorem 2, we know that regardless of the relative position of the known Bass diffusion process in the corresponding VBM, the same VBM is obtained by backward transforming the known process to $-\infty$; thus, depending on the start time of the known process, the original parameters are different, but the VBM's $M \equiv (1+p/q)m$ is a constant. In our forward and backward transformations, the sum of $p$ and $q$ does not change in any transformation; thus, the VBM's $U$ is also a constant.

[2] Strictly speaking, the parameter values in Eqs. (18), (19), (20) should be expressed as limits. For example, $p'=0$ should be $\lim_{\tau\to\infty} p'=0$, implying that as the assumed start time of a diffusion process is extended to negative infinity, the value of $p'$ becomes infinitely close to zero; however, this result does not imply that we can start from time negative infinity with $p=0$, $q=U>0$, and $m=M>0$ and then derive other results by forward transformation.

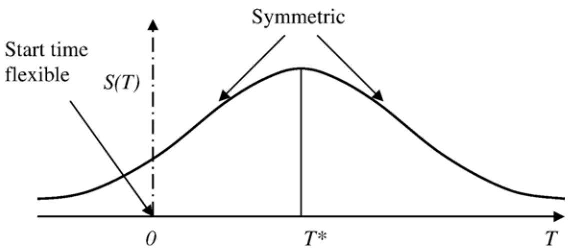


Fig. 5. The Virtual Bass Model.

(II) We now check the diffusion process that starts from the other end, positive infinity. By letting $\tau$ go to $\infty$ in the forward transformation, we have:

$$Y(\infty) = m,\quad \begin{cases} m' = m - Y(\infty) = 0, \\ q' = q[m - Y(\infty)]/m = 0, \\ p' = p + (q/m)Y(\infty) = p + q. \end{cases} \tag{20}$$

We see that for the diffusion process that starts from positive infinity, $m'$ equals zero, and the values of $p'$ and $q'$ are switched from the negative infinity case.

(III) For a diffusion process that starts from the VBM's peak time, we obtain the three parameters by forward transformation if $p \le q$ or by backward transformation if $p > q$:

$$\begin{cases} m' = m(1 + p/q)/2 = M/2, \\ q' = p' = (p + q)/2 = U/2. \end{cases} \tag{21}$$

Now we have a clearer picture of how the parameters change as the start time of a VBM process moves from $-\infty$ to $+\infty$: $m$ decreases from $M$ to 0, $q$ decreases from $U$ to 0, and $p$ increases from zero to $U$. As the process moves, the sum of $p$ and $q$ is constant, which implies that some *value* is continuously shifting from $q$ to $p$. For a start time at $T^*$, $m = M/2$, and $p = q = U/2$.

### 3.4. Virtual Bass Model equations

We have shown that two of the constant parameters $M$ and $U$ of a VBM can be obtained by Eqs. (V1) and (V2), defined below. We now examine the other parameters in the model. To permit time to vary according to start times, we denote the start time relative to the time of peak of a Bass diffusion process by $z$. We set $z = 0$ for the process that starts from the peak time point of the VBM curve; therefore, a Bass diffusion process that starts before the peak point has a negative $z$ value and a process that starts after the peak point has a positive $z$ value. For a given process, if $p$ and $q$ are given, $z$ can be calculated by Eq. (V3). From Eq. (V3) we see that $z$ takes a negative value if $p < q$. We denote the Bass diffusion process that starts from time $z$ by $D(z)$ and its diffusion rate function by $S_{(z)}(T)$.[3] For example, $S_{(0)}(T)$ represents the diffusion rate of the diffusion process with $z = 0$. For such a diffusion process, the peak diffusion rate, $H$, is a constant along with $M$ and $U$ and is given by:

$$H = S_{(0)}(0) = (M/2)(U/2) = MU/4.$$

The *Virtual Bass Model Equations* (VBE), shown below, summarize the relationships among the parameters. By examining (VBE), we conclude: the parameters ($z'$, $q'$, $p'$, and $m'$) for

$$\begin{cases} M = (1 + p/q)m & \text{(V1)} \\ U + p + q & \text{(V2)} \\ z = [1/(p + q)]\ln(p/q) & \text{(V3)} \\ z' = z + \tau; \quad q' = U/(1 + e^{z'U}); \quad p' = U - q'; \quad m' = M/(1 + p'/q'); & \text{(VBM)} \end{cases} \tag{VBE}$$

a transformed diffusion process that starts $\tau$ periods away from the start time of the original process is obtained from the four equations (Eq. (VBE)). The transformed parameters provide us with a unique $D_{(z')}$. Eq. (VBE) enables us to compute the parameters for a new process from the parameters of the known process without explicitly going through the forward or backward transformations shown in Subsections 3.1 and 3.2 Eq. (VBE) may be used for both forward and backward transformations and for the cases of $p \le q$ as well as $p > q$.

### 3.5. An example of backward transformation

To illustrate the application of VBE we use the case of color television for the 1963–1970 data series from Mahajan, Mason, and Srinivasan (1986). The parameters estimated by the Nonlinear Least Squares method (Srinivasan and Mason, 1986) are: $m = 39658.62$, $p = 0.018466$, and $q = 0.615863$. The first sales of color television began in 1954 and thus the parameter estimates based on a start time of 1963 suffer from a left-hand data-truncation bias.

We use the VBM transformations from the Eq. (VBE) with $\tau = -9$. We first calculate the two constants based on Eqs. (V1) and (V2):

$$M = (1 + p/q)m = 40,847.74, \text{ and } U = p + q = 0.634329.$$

The parameter $z$ for the process starting from 1963 is calculated from Eq. (V3) to be $z = -5.52882$; thus, $z'$ for the process starting from 1954 is: $-14.52882$. Using the other VBE equations we have:

$$q' = U/(1 + e^{z'U}) = 0.634265935,$$
$$p' = (U - q') = 6.3065E{-}5, \text{ and}$$
$$m' = M/(1 + p'/q') = 40,843.68.$$

The calculated value of $T^* = \frac{1}{(p+q)}\ln\left(\frac{q}{p}\right)$ is 5.5 for the first series and 14.5 ($T'^* = \tau + T^*$) for the series beginning in 1954, and the predicted sales rate at the peak is 6477 (in thousands) in both cases.

The VBM procedure for obtaining adjusted parameter values just illustrated is an *indirect* approach in that the parameters $m$, $p$, and $q$ are estimated from the original data series and then transformed by the VBM equations. There does exist, however, a *direct* method to estimation that will produce equivalent

[3] Although the values of $m$, $p$ and $q$ in Eqs. (V1) (V2) (V3) are different with different $z$, we have chosen not to denote them by $p(z)$, $q(z)$ and $m(z)$ for notational simplicity and clarity.

parameter estimates for $m'$, $p'$, and $q'$. In this approach we consider two frames for time. The periods of the data are numbered starting at time $T=1$. If the data are left-truncated, the product introduction time is earlier and we will refer to its timeframe as "absolute" time starting with 1 as the period in which the product was introduced. If $\tau$ is the difference between the product introduction and the data start, the time of the data *relative* to the product introduction is $\tau+\mathrm{T}$, which starts at $\tau+1$. The direct method expression for Eq. (4) becomes

$$Y(T)=\frac{m\left(1-e^{-(p+q)(\tau+T)}\right)}{(q/p)e^{-(p+q)(\tau+T)}+1}. \qquad (22)$$

The parameter values resulting from the estimation of $Y(T)-Y(T-1)$ using Eq. (22) in the Srinivasan and Mason (1986) form of the Bass model will be the same (up to several decimal places) as those obtained by the VBM procedure. For example, using the color television data, the initial value for time is set equal to 10 (representing 1963) and the next value is set equal to 11 (representing 1964) and so on. If nonlinear least squares regression is employed, the resulting parameter values will differ only to a trivial degree from the VBM method. To demonstrate this we show below a comparison of the estimates in Table 4.

The values for $p'$ differ only in 6 decimal places and for $q'$ differ only in 8 decimal places, while the estimates of $m'$ are also very close to one another. These results are typical of the many comparisons that we have made using the VBM method and the direct method. In every case that we have examined the estimates $p'$ and $q'$ differ only in several decimal places. We therefore believe that the two methods are "equivalent," although there are very small differences between them. We think that these differences are likely due to imperfect machine precisions and that the two methods are mathematically the same. We present below an argument that leads us to believe the two methods are mathematically the same.

As indicated in Fig. 6, when the VBM approach is used, the start time is $T_1$ and the estimation takes place over the interval from $T_1$ to $T_2$. The parameter estimates are then transformed on the basis of information about the true introduction time, $T_0$. In the direct approach the estimation is also over the interval between $T_1$ and $T_2$, but the initial value for time will be $\tau+1$ rather than 1 used as the initial value in the indirect approach. Both approaches will have the same fit to the data and both will produce the same fitted curve over the interval covered by the data (Note that the error structure of the data is the same in both cases). From Theorem 2 we know that any segment of a Bass diffusion curve uniquely determines the backward and forward transformations. Thus both curves will match between time $T_0$ and $T_1$; therefore, *when nonlinear least squares is used with time as the exogenous variable the direct approach and the VBM approach to estimation for the true introduction time are equivalent*. This conclusion has been confirmed by substantial experimentation.

Table 4
Comparison of VBM and direct parameter estimates

| Parameter | VBM | Direct | Difference |
|---|---|---|---|
| $q'$ | 63465935 | 0.634267 | −1.06617E−06 |
| $p'$ | 6.3065E−05 | 6.3066E−05 | 6.61657E−08 |
| $m'$ | 40843.68 | 40843.71 | −0.029426755 |

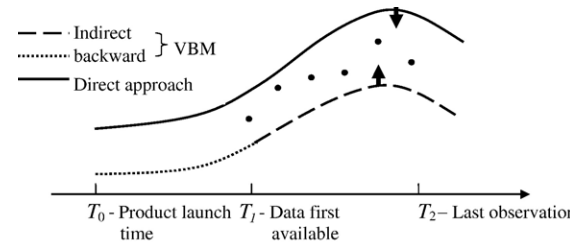


Fig. 6. Direct approach and VBM approach with left-truncated data.

VBM provides a mathematical justification for the use of the (previously intuition-based) direct approach to dealing with left-hand truncation bias. Our theoretical developments here provide the flexibility of adjusting for the left-hand truncation bias when the data are not available but when the parameters have been estimated and reported. These parameter estimates may be transformed by VBM to be consistent with different introduction times.

## 4. Evaluation of other methods of estimating with left-truncated data

Here we consider three other possible methods of estimating with left-truncated data. Lilien et al. (2000) suggested a method for dealing with left truncation in which penetration data are used for estimation along with the discrete version of the differential equation of the Bass model, which is time invariant. Because the Bass model differential equation is a function of cumulative adoptions, estimates of parameters are theoretically equal for left-truncated and non-left-truncated series and offer the possibility, which we will explore, of being substituted into the Bass model differential equation solution for forecasting as a function of time and calculation of the time of peak sales. A reviewer suggested another possible way to adjust for left truncation, which we investigate, by using the Mahajan and Peterson (1985) solution to the Bass model differential equation, which has the constant of integration set at $t=t_0$, so that $Y(t_0)=Y_0$, where $Y_0$ is cumulative adoption at $t_0$, rather than the usual condition of at $t=0$, so that $Y(0)=0$. We have also experimented with a third method: estimating $\tau$, the start time of the data relative to product introduction, in the Bass model direct method.

### *4.1. Use of the discrete version of the differential equation and penetration data*

Lilien et al. (2000) have proposed that to deal with the left-truncation issue the discrete version of the Bass model differential equation could be used together with penetration data of household ownership. The discrete version of the differential equation is:

$$x(t)=pm+(q-p)X(t-1)-(q/m)[X(t-1)]^2, \qquad (23)$$

where $x(t)$ is the number of adoptions occurring in period $t$, $X(t-1)$ is the cumulative number of adoptions having occurred before period $t$, and $m$, $p$, and $q$ are the Bass model parameters. Eq. (23) depends only on cumulative adoptions and not on time. It therefore appears that this equation estimated on household penetration data offers a viable approach to estimation with data that are left-truncated. Lilien et al. used change in penetration for $x(t)$ and estimated Eq. (23) with nonlinear least squares and non-negativity constraints on the parameters for several data series. In some instances the estimated value of $p$ was 0, but in many cases their estimates appear to be plausible. To compare the estimation with the discrete version of the differential equation as suggested by Lilien et al. with the direct method we have conducted a simulation analysis.

### 4.2. Simulation analysis of use of discrete time differential equation

We simulated a set of 20 data series with the Bass model. In the simulations we chose parameters to reflect those found in typical diffusion data. These chosen parameter values imply different ranges of the time from introduction to peak sales. Some of the data series have long left tails of more than 20 years to the peak, others have a medium left tails with ranges of 11 to 18 years, and two still others have short left tails with 10 or fewer years. We have simulated the data using the Srinivasan and Mason (1986) method of taking first differences of $mF(t)$ and adding error. We use this approach because the real data are discrete and this seems to be the best way to represent differenced penetration data. In 12 of the simulated series we used homoskedastic error and in 8 cases we used heteroskedastic error as a means for increasing error with observations. To study the effects of differing degrees of left truncation, we estimated with an early start period and a late start period for the data for series. These start times are shown in Table 5.

For each data series we estimated the parameters with the differential equation indicated in Eq. (23) and with the Srinivasan–Mason method. We estimated with starting points for the data: (a) from the early start time of the data, and (b) for the late start time of the data. In most instances the ending data period used in the estimation was the first data period past the peak. In a few instances where the peak was "flat" we included a few additional data points in order to be consistent with the rule we used in estimating empirical data.

Shown in Table 5 below are: the true parameters used in the simulations and estimates of the actual and predicted time of peak sales derived from the parameter estimates for the differential equation and for the direct method for the early starting point and the late starting point. We used the equation: $\left(\frac{1}{p+q}\right)\ln\left(\frac{q}{p}\right)+1$ to obtain the predicted time of peak, $T^*$, for the differential equation estimates and the direct method estimates. The +1 was added to the formula in Eq. (5) because (5) is based on a continuous time assumption that starts at 0, while the data and estimations are based on discrete time for the direct estimation with data starting at $\tau+1$, where $\tau$ is the difference between the start period of the estimation and the time of the first data point. As indicated by Eq. (23) the differential equation is a discrete equation in which time is not an explicit variable. The estimates for the direct method were developed using the Srinivasan and Mason (1986) method of first differences in the cumulative version of the Bass model ($m[F(t)-F(t-1)]$). The estimates of the differential equation were developed using nonlinear least squares. In both cases bounds were placed on the parameter estimates requiring non-negativity.

Table 5
Comparison of estimates of time to peak adoption, $t*$, for differential equation and direct methods with different start periods for data (with $Y_0$ given for differential equation)

| Parameters for simulated non-truncated data | | | | Truncation level of simulated data | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Early data start period | | | Late data start period | | |
| | | | | Data start | $t$* estimates | | Data start | $t$* estimates | |
| See note | $p$ | $q$ | Actual $t$* | | Dif. eqn. | Direct method | | Dif. eqn. | Direct method |
| | 7.70E −07 | 0.44 | 31 | 10 | 13 | 30 | 20 | 11 | 30 |
| | 1.00E −05 | 0.4 | 27 | 10 | 14 | 27 | 20 | 11 | 27 |
| # | 5.00E −05 | 0.35 | 26 | 6 | 16 | 26 | 12 | 15 | 26 |
| | 5.00E −05 | 0.48 | 23 | 10 | 11 | 23 | 20 | 7 | 23 |
| # | 0.001 | 0.3 | 20 | 6 | Bound $p$ | 19 | 12 | Bound $p$ | 19 |
| # | 0.005 | 0.195 | 20 | 4 | 17 | 18 | 10 | Bound $p$ | 18 |
| | 0.001 | 0.25 | 20 | 4 | Bound $p$ | 20 | 10 | Bound $p$ | 20 |
| # | 6.00E −06 | 0.9 | 20 | 6 | 9 | 20 | 12 | 7 | 20 |
| # | 5.00E −05 | 0.4 | 20 | 6 | 16 | 21 | 12 | 15 | 21 |
| # | 0.008 | 0.145 | 18 | 7 | 18 | 19 | 11 | Bound $p$ | 19 |
| | 1.00E −04 | 0.6 | 15 | 5 | 8 | 15 | 9 | 7 | 15 |
| | 9.00E −05 | 0.63 | 15 | 5 | 8 | 15 | 10 | 6 | 15 |
| | 0.02 | 0.08 | 14 | 4 | 13 | 14 | 10 | Bound $p$ | 13 |
| # | 0.005 | 0.3 | 12 | 4 | 12 | 14 | 7 | 12 | 14 |
| | 0.02 | 0.22 | 11 | 4 | 9 | 10 | 6 | 8 | 11 |
| # | 0.01 | 0.3 | 10 | 4 | 10 | 12 | 7 | 9 | 12 |
| | 0.01 | 0.4 | 9 | 3 | 7 | 9 | 5 | 9 | 9 |
| | 0.02 | 0.34 | 9 | 3 | 6 | 8 | 5 | 8 | 8 |
| | 0.01 | 0.5 | 8 | 3 | 5 | 8 | 5 | 7 | 8 |
| | 0.035 | 0.4 | 5 | 3 | 4 | 6 | 5 | 4 | 5 |

Note: # indicates that the disturbances were simulated as heteroskedastic with the standard deviation proportional to adoption. All other simulations were made under the assumption of homoskedastic disturbances.

As indicated in Table 5, in most cases the direct method produced estimates of the time of peak sales that is the same or one period away from the true peak time of the simulated series and in no case did it differ by more than two periods from the true time of peak sales. On the other hand the differential equation produced estimates that were unsatisfactory in every case in which the time to peak sales was greater than 14 years either because the estimate of p was at the lower bound of 0 or because the parameter estimates implied a time of peak sales that was far away from the true time of the peak. Only in the absence of a long left tail does the differential equation sometimes produce parameter estimates that are close to the true values and provide good predictions of the time of peak sales. Since the estimates of $T^*$ from the differential equation are far from the true time to peak in most cases, we believe that the parameter estimates based on the differential equation will ordinarily not be useful in predicting the time to peak or in forecasting future sales using any Bass model adoption equation with time as an argument.

In summary, the direct method or VBM is superior to estimation with the discrete version of the differential equation. Although theoretically the differential equation method is time invariant this desirable feature must be set against the weaknesses of its estimation properties. Putsis and Srinivasan (2000), in a review paper on estimation methodology for diffusion models, concluded that estimation of the discrete version of the differential equation of the Bass model with OLS regression may produce parameter estimates that are extremely unstable in the presence of few data points and often yield parameter estimates that have the wrong sign. The differential equation cannot be counted on to produce parameter estimates that can be used in the differential equation solution to calculate sales as a function of time. Parameter estimates using the differential equation, therefore, cannot be used to determine the extent of left truncation or to correct for it. Although we do not report the details here, we have conducted simulations with multiple replications of simulated data based on parameters and standard deviations estimated from actual data for several products. We confirmed that estimations using the differential equation do not adequately recover the parameters used in the Srinivasan–Mason form of the Bass model used to generate the data.

### 4.3. Does the Mahajan–Peterson equation adjust for left truncation?

The differential equation for the Bass model is:

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = pm + (q-p)Y - \left(\frac{q}{m}\right)Y^2, \tag{24}$$

where: $Y=mF(t)$ is cumulative adoption at $t$. The general solution to this equation is:

$$Y(t) = \frac{\left(m - pe^{-(p+q)(t+C)}\right)}{\left(1+\left(\frac{q}{m}\right)e^{-(p+q)(t+C)}\right)}.^{4} \tag{25}$$

If $C$ is determined so that $Y(0)=0\left[C=-\frac{1}{p+q}\ln\left(\frac{m}{p}\right)\right]$, the solution for $Y(t)$ will be that given by Eq. (4) and the resulting adoption function will be that indicated in Eq. (3). But, as shown by Mahajan and Peterson (1985), if the condition that $Y(t=t_0)=Y_0$ is imposed in determining the constant of integration the solution will be:

$$Y(t) = \frac{\left(m - \frac{p(m-Y_0)e^{-(p+q)(t-t_0)}}{\left(p+\left(\frac{q}{m}\right)Y_0\right)}\right)}{\left(1+\frac{\left(\frac{q}{m}\right)(m-Y_0)e^{-(p+q)(t-t_0)}}{\left(p+\left(\frac{q}{m}\right)Y_0\right)}\right)}, \tag{26}$$

and the resulting adoption equation will be:

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \frac{\left(\frac{(p+q)^2(m-Y_0)e^{-(p+q)(t-t_0)}}{\left(p+\frac{q}{m}Y_0\right)}\right)}{\left(1+\frac{\frac{q}{m}(m-Y_0)e^{-(p+q)(t-t_0)}}{p+\frac{q}{m}Y_0}\right)^2}. \tag{27}$$

Although the Mahajan–Peterson equation does not impose the condition that at $t=0$, $Y(0)=0$ as does the Bass model, in the real world there can be no sales prior to the product introduction. The Mahajan–Peterson equation is logically inconsistent with this real-world condition unless $Y_0$ is known precisely and without error such that $Y(0)=0$ is also true. Because of this logical inconsistency, projected early data points using the Mahajan–Peterson equation can be poor and even negative, in some cases.

In Eqs. (26) and (27) the parameter $t_0$ is not intended to relocate the equation in time relative to the product introduction, which would adjust for left truncation. Rather, as illustrated in Mahajan (1975), the expression $(t-t_0)$ adjusts an absolute time value (e.g., year) to start at 0 in Eq. (26) and 1 in Eq. (27); for example, with data starting at product introduction in 1963 using Eq. (26), $(t-t_0)$ would start at $(t-t_0)$=(1963–1963)=0, followed by (1964–1963)=1, then (1965–1963)=2, etc.. Although it is possible to modify the time expression to adjust for left truncation, thus creating a direct method for the Mahajan–Peterson equations similar to that of the Bass model in Eq. (22), it offers no advantages compared to the Bass model direct method and has the disadvantageous logical inconsistency described in the preceding paragraph.

### 4.4. Can the time the data start relative to the product introduction time be estimated using the Bass model direct method?

In Eq. (22), the Bass model direct method for left-truncation adjustment, $\tau$ is the start time of the data relative to product introduction. If it were possible to treat $\tau$ as a parameter to be estimated, then left-truncation adjustment could be accomplished without knowing the product introduction date or the start time of the data relative to it. Our preliminary experiments involving simulated data with very small error suggest that this might be possible in idealized cases; however, as error increases, estimating $\tau$ becomes increasingly problematic. Since in the real world the error magnitude is not known, there is no way to know if an estimate of $\tau$ is plausible without knowing what it should be in which case there is no need to estimate it. Nonetheless, we

[4] There are multiple solutions to the Bass model differential equation each with its unique constant of integration. After substitution of the corresponding constant of integration so that $Y(0)=0$, the solutions are equivalent.

believe that knowledge of the Bass model will be advanced by further research with regard to estimating $\tau$, but we do not expect that $\tau$ estimation will ever be preferred to determining $\tau$ through product background research and using the direct method or VBM to adjust for left-truncated data.

## 5. Correcting left-hand truncation bias for use in guessing by analogy

We examine here the extent of the left-hand truncation bias issue for new products, services, and technologies. To do this we determine the introduction time for a substantial number of new products from historical sources and collect the available sales data for these products. We can then measure the extent of the problem and employ the VBM methodology to correct the parameter estimates where the data start time is different from the product introduction time. In this way we will create a database of appropriate $p'$s and $q'$s for use in guessing by analogy.

### *5.1. The introduction time issue and data collection*

Considered in the context of guessing by analogy, the determination of the introduction time to be used is not a simple matter. For a manager attempting to make a forecast for a new product prior to product launch and seeking an appropriate analogy, judgments must be made about the state of development of the new product, the state of available infrastructure support, the level of marketing support, and other considerations that will influence whether or not the new product sales will takeoff early or whether sales will initially grow slowly for a period of time before takeoff. The manager will want the introduction time to be analogous to that provided by historical information for new products. In some cases the introduction time will be defined as the time in which the product was first made available for sale, but under other circumstances the introduction time may be judged to be the year in which significant sales begin.[5] The matter is further complicated by the fact that the commercialization date indicated by historical sources may refer to the time when a business unit was created or when a patent was granted and will not be the time of first product sales. In some cases historical sources will differ in introduction dates. We have applied our judgment to resolve uncertain or conflicting information in deciding on the introduction years used in this study.

We consulted numerous historical sources for information about the introduction year for the products in this study. These historical sources as well as sources used for the products sales or subscriptions are listed in Appendix B. We excluded products for which we could not obtain the introduction date. In total we have constructed a database of 39 products and services in five different categories: home appliances (10 products), housewares (7 products), consumer electronics (11 products), products bought by both business and consumers (7 products), and subscription services (4 products). Products in the first three categories have been included in previous studies of comparative diffusion rates, while the last two categories have not.

### *5.2. Parameter estimates and time to peak sales comparisons*

For the 39 products in our study we have estimated the Bass model parameters from the start time of the data and from the year of introduction using the method described in Srinivasan and Mason (1986) for nonlinear least squares. As indicated by Putsis and Srinivasan (2000) this method has become the de facto standard in diffusion research. In those cases where the start time of the data differs from the year of introduction we have used the VBM method to obtain parameter estimates that are appropriate for the year of introduction. The data interval used for estimation included the range from the start of the data to one period past the peak in sales, in most cases. The exceptions included those cases where the last observation of the data was treated as the peak and those cases where the sales at the peak and at one period past the peak were very close to one another. In no instance did the data include more than two periods past the peak. Our rationale for the data range is that we want to capture the curvature in the data, but we also want to exclude data points that are increasingly contaminated by repeat purchases.

Shown in Table 6 (below) are the nonlinear least squares estimates for the Bass model $p$ and $q$ parameters for the 39 products in our analysis (as indicated by $p$ data and $q$ data) over the time periods for which data are available for each product (indicated under the column Data Used). Also shown are the VBM estimates based on the year each product was introduced (as indicated by $p$ VBM and $q$ VBM). The actual years from introduction to peak sales are indicated by $T^*$ act (actual) and the predicted years to peak sales based on the VBM estimates of $p'$ and $q'$ and the predicted years to peak sales based on the parameter estimates of $p$ data and $q$ data.[6] In all but 7 cases the product introduction year is different from the first year of available data; thus, we conclude that the extent of the left-hand truncation bias is substantial. In most cases there is a large difference between the actual years to peak sales and the predicted years to peak sales based on parameter estimates derived from the data available. Thus reliance on the estimates of $p'$s and $q'$s from most previously published databases for guessing by analogy can be strongly questioned. The prevalence of long left tails in the diffusion process is greater than would be suggested by previous studies and we believe that managers would be well advised to take this into account when forecasting prior to product launch. Table 6 provides the correct parameter values adjusted for the introduction time of the product.

Shown in Fig. 7 is the cumulative distribution of the number of years to peak sales for the 39 products in our study. Only a single product reached peak sales within 5 years and for 90% of the products it took 10 years or more for sales to reach the peak

[5] For an example of the use of an analogy with a start time different from the time of introduction see Bass et al. (2001) in which cable television in the 1980's was used to calibrate the diffusion of satellite television.

[6] The predicted time of peak for the VBM is $\frac{1}{p+q}\ln\left(\frac{q}{p}\right)+1$. The +1 was added to the formula in Eq. (5) because the assumption here is one of discrete time starting at $\tau$+1 rather than continuous time starting at 0.

Table 6
Comparison of time from introduction to peak sales, VBM estimates and estimates based on data

| Categories | Year introduced | $t*$ act | $t*$ VBM | $t*$ data | $p$ VBM | $q$ VBM | $p$ data | $q$ data | Data used |
|---|---|---|---|---|---|---|---|---|---|
| Home appliances | | | | | | | | | |
| Clothes dryers | 1930 | 27 | 27 | 7 | 1.4E−06 | 0.4792 | 0.0199 | 0.4593 | 1950–1957 |
| Clothes washers | 1910 | 20 | 19 | 7 | 0.00162 | 0.2687 | 0.03623 | 0.234 | 1922–1930 |
| Electric range | 1919 | 12 | 11 | 5 | 0.00246 | 0.4984 | 0.04543 | 0.4554 | 1925–1931 |
| Freezers | 1929 | 25 | 25 | 8 | 3.8E−05 | 0.3813 | 0.02359 | 0.3578 | 1946–1954 |
| Microwave ovens | 1955 | 33 | 33 | 18 | 4E−06 | 0.3451 | 0.00071 | 0.3444 | 1970–1988 |
| Power lawnmowers | 1926 | 34 | 35 | 13 | 7.9E−06 | 0.3091 | 0.00691 | 0.3022 | 1948–1960 |
| Refrigerators | 1913 | 28 | 28 | 21 | 0.00037 | 0.2308 | 0.00188 | 0.2293 | 1920–1940 |
| Room air conditioners | 1928 | 29 | 29 | 11 | 4.4E−08 | 0.5701 | 0.00125 | 0.5689 | 1946–1957 |
| Trash compactors | 1964 | 11 | 11 | 4 | 6.5E−05 | 0.9498 | 0.04766 | 0.9023 | 1971–1975 |
| Vacuum cleaners | 1908 | 22 | 21 | 7 | 0.00406 | 0.1805 | 0.04238 | 0.1422 | 1922–1930 |
| Housewares | | | | | | | | | |
| Blenders | 1946 | 24 | 25 | 16 | 3.8E−06 | 0.4726 | 0.00027 | 0.4724 | 1955–1970 |
| Broilers | 1937 | 18 | 18 | 9 | 3.6E−08 | 0.9668 | 0.00022 | 0.9667 | 1946–1956 |
| Coffee makers | 1934 | 23 | 23 | 9 | 3.4E−05 | 0.4086 | 0.01023 | 0.3984 | 1948–1957 |
| Heating pads | 1918 | 12 | 14 | 10 | 0.0035 | 0.3463 | 0.01375 | 0.3361 | 1922–1930 |
| Electric blankets | 1930 | 32 | 34 | 15 | 5.7E−05 | 0.2489 | 0.00631 | 0.2427 | 1949–1961 |
| Electric shavers | 1931 | 26 | 29 | 12 | 9.8E−05 | 0.2775 | 0.01057 | 0.2670 | 1948–1958 |
| Steam irons | 1936 | 21 | 22 | 9 | 0.00012 | 0.3819 | 0.01693 | 0.3651 | 1949–1957 |
| Consumer electronics | | | | | | | | | |
| B&W TV | 1939 | 17 | 16 | 9 | 0.00064 | 0.416 | 0.01156 | 0.4051 | 1946–1956 |
| Camcorders | 1973 | 18 | 19 | 7 | 9.4E−05 | 0.4679 | 0.02441 | 0.4436 | 1985–1991 |
| Cassette decks | 1964 | 20 | 20 | 10 | 0.001 | 0.2875 | 0.1688 | 0.2717 | 1974–1984 |
| CD players | 1983 | 13 | | 14 | | | 0.00170 | 0.3991 | 1983–1996 |
| Color TV | 1954 | 15 | | 15 | | | 0.00005 | 0.6480 | 1954–1969 |
| Digital watches | 1971 | 13 | 12 | 9 | 0.0056 | 0.3542 | 0.01652 | 0.3542 | 1974–1984 |
| Laser disc players | 1980 | 14 | 15 | 10 | 0.0025 | 0.3242 | 0.01243 | 0.3143 | 1985–1995 |
| Projection TV | 1984 | 16 | | 18 | | | 0.00512 | 0.2062 | 1984–2001 |
| Radio | 1922 | 8 | | 9 | | | 0.01034 | 0.4537 | 1922–1931 |
| Record players | 1906 | 55 | 53 | 7 | 1.8E−11 | 0.4519 | 0.01870 | 0.4332 | 1952–1961 |
| VCR's | 1975 | 11 | 12 | 7 | 0.00015 | 0.7564 | 0.00637 | 0.7501 | 1980–1987 |
| Business and consumer products | | | | | | | | | |
| Answering machines | 1960 | 30 | 30 | 8 | 5.8E−08 | 0.5433 | 0.00885 | 0.5345 | 1982–1990 |
| ATM machines | 1971 | 14 | | 14 | | | 0.00053 | 0.4957 | 1971–1985 |
| Copying machines | 1960 | 20 | 21 | 16 | 0.00573 | 0.1519 | 0.01208 | 0.1456 | 1965–1981 |
| Fax machines | 1980 | 18 | 19 | 12 | 0.00198 | 0.2637 | 0.01220 | 0.2535 | 1987–1998 |
| Handheld calculators | 1967 | 14 | 14 | 11 | 0.00136 | 0.4380 | 0.00505 | 0.4343 | 1970–1981 |
| PC printers | 1976 | 9 | 9 | 5 | 0.00071 | 0.8037 | 0.01743 | 0.7870 | 1980–1986 |
| Portable dictation machines | 1955 | 25 | 27 | 12 | 0.00078 | 0.2141 | 0.00650 | 0.2084 | 1965–1981 |
| Subscription services | | | | | | | | | |
| AOL change in subs. | 1989 | 8 | 8 | 4 | 0.00018 | 1.1827 | 0.02051 | 1.1624 | 1933–1997 |
| Cable TV change in subs. | 1948 | 23 | 23 | 18 | 6.1E−06 | 0.5012 | 0.00001 | 0.5013 | 1953–1971 |
| Cell phones (analog) change in subs. | 1983[a] | 13 | | 16 | | | 0.00074 | 0.4132 | 1984–1995 |
| Satellite TV change in subs. | 1994 | 5 | | 6 | | | 0.04693 | 0.3346 | 1994–1998 |

[a] Partial year.

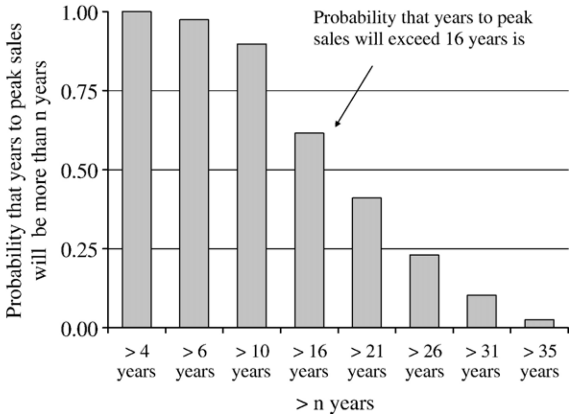


Fig. 7. Cumulative probability of years to peak sales.

from the year of introduction. On average it took 20 years from the time of introduction for the sales to reach peak sales.

In Table 7 the averages for each of the measures in Table 6 are shown by each of the product categories. These averages do not vary a great deal over the categories although consumer electronics has the shortest mean time to peak sales at 18 years and home appliances has the longest mean time to peak at 24 years. Overall, the mean time to peak sales is 20 years. It is noteworthy that the overall average of predicted time to peak sales based on the start time of the data is less than 11 years and, in comparison with the actual average time to peak sales from product launch of 20 years, indicates a large left-truncation bias.

## 6. Conclusions

The use of guessing by analogy in conjunction with the Bass model for forecasting the diffusion of new products and technologies prior to product launch has brought to the fore the importance of the left-hand truncation bias. We have demonstrated here that for most of the products for which there is historical information about the time of introduction the data series for sales of these products do not extend back to the introduction date. In these cases the parameter estimates of the Bass model will be biased and will imply a shorter length of time for sales to reach the peak than the actual length. We have shown that the average time to peak over the 39 products for which we have information about the introduction time is 20 years. This is almost double the average time to peak when measured from the start time of the data. The prevalence of long left tails in the diffusion patterns for new innovations is a notable finding of this study.

Managers relying on databases of Bass model parameter estimates in which the estimates have not been adjusted for the left-hand truncation bias can be badly misled. We have shown that the Virtual Bass Model may be used to transform biased parameter estimates so that the left-truncation bias is removed. We have provided in Table 6 a database of $p'$s and $q'$s ($p$ VBM and $q$ VBM) that can be used in guessing by analogy. We have also demonstrated that if the introduction time of a new product is known it is possible to obtain the adjusted parameter estimates easily by use of the direct method in which the start time of the regression is set equal to the time difference between the start of the data and the introduction time plus 1.

In creating the Virtual Bass Model we have uncovered previously unknown mathematical properties of the Bass model. We have proven that for any Bass diffusion process with a given start time there is an equivalent diffusion process for any other start time. We have also proved that the transformation of parameters in moving from one start time to another is unique. The Virtual Bass Model has both theoretical and practical significance. On the practical side it can be used to transform parameter estimates based on left-hand truncated data into unbiased parameter estimates. We also believe that through increased understanding of the properties of the Bass model the theoretical development we have presented here stands a good chance of finding other practical applications of significance.

## Acknowledgements

The authors would like to thank three anonymous reviewers for constructive suggestions and ideas that have improved this paper. We also wish to thank the editor, Hubert Gatignon, for his wisdom and patience in guiding us through the various revisions of the paper.

## Appendix A. Proof of Theorem 2

Suppose we have a Bass diffusion curve segment between time $A$ and $B$. To prove that the curve segment uniquely determines a VBM curve, we only need to show that for any diffusion start time $z<A$, there is only one unique diffusion curve that begins from $z$ and matches the known curve segment completely.

Suppose that we have found such a diffusion curve $S(T)$. If $S(T)$ is not the unique solution, there must exist another $S'(T)$ that also starts from $z$ and matches the same curve segment. Then the two curves $S(T)$ and $S'(T)$ must overlap between $A$

Table 7
Averages by category of actual time from introduction to peak sales and comparison of peak time and parameters of VBM and data estimates

| | $t^*$ actual | $t^*$ VBM | $t^*$ data | $p$ VBM | $q$ VBM | $p$ data | $q$ data |
|---|---|---|---|---|---|---|---|
| Home appliances | 24.1 | 23.9 | 10.1 | 0.0009 | 0.4213 | 0.0226 | 0.3996 |
| Housewares | 22.286 | 23.571 | 11.429 | 0.0005 | 0.4432 | 0.0083 | 0.4355 |
| Consumer electronics | 18.182 | 21 | 10.455 | 0.0014 | 0.4369 | 0.0113 | 0.4254 |
| Business & consumer products | 18.571 | 20 | 11.143 | 0.0018 | 0.4025 | 0.009 | 0.4084 |
| Subscription services | 12.25 | 15.5 | 11 | 1E−04 | 0.842 | 0.17 | 0.6029 |
| Overall averages | 19.897 | 21.462 | 10.718 | 0.001 | 0.4694 | 0.0138 | 0.4357 |

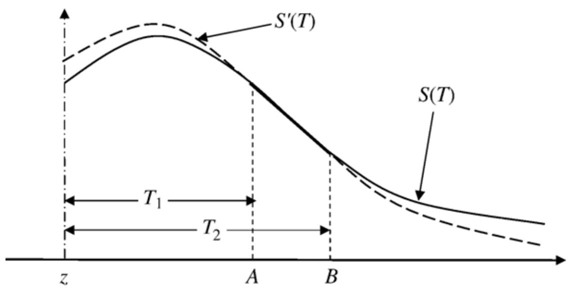


Fig. A1. If two VBM curves both fit a Bass diffusion curve segment….

and $B$. The scenario is shown in Fig. A1. We denote the three parameters for $S(T)$ by $m$, $p$, $q$, and those for $S'(T)$ by $I'$, $p'$, and $q'$. Since $S(T)$ and $S'(T)$ are different, at least one of their parameters must be different. We denote the distance between $z$ and $A$ by $T_1$ and the distance between $z$ and $B$ by $T_2$.

Eqs. (A1), (A2), and (A3) hold since $S(T)$ and $S'(T)$ overlap between time $A$ and $B$.

($\tau$ satisfies $0 \le \tau \le (T_2 - T_1)$ throughout the proof.)

$$S(T_1) = S'(T_1) \tag{A1}$$

$$S(T_1+\tau) = S'(T_1+\tau) \tag{A2}$$

$$Y(T_1+\tau)-Y(T_1) = Y'(T_1+\tau)-Y'(T_1) = I(\tau). \tag{A3}$$

$I(\tau)$ in Eq. (A3) represents the number of new adoptions from time $A$ to $(A+\tau)$. From Eq. (A3), we have:

$$\begin{aligned} S(T_1+\tau)-S(T_1) &= pm + (q-p)Y(T_1+\tau)-(q/m)[Y(T_1+\tau)]^2 \\ &\quad -(pm + (q-p)Y(T_1)-(q/m)[Y(T_1)]^2) \\ &= (q-p)[Y(T_1)+I(\tau)]-(q/m)[Y(T_1)+I(\tau)]^2 \\ &\quad -(q-p)Y(T_1) + (q/m)[Y(T_1)]^2 \\ &= [(q-p)-2(q/m)Y(T_1)]I(\tau)-(q/m)[I(\tau)]^2. \end{aligned} \tag{A4}$$

Similarly, we have Eq. (A5):

$$\begin{aligned} &S'(T_1+\tau)-S'(T_1) \\ &= [(q'-p')-2(q'/m')Y'(T_1)]I(\tau)-(q'/m')[I(\tau)]^2. \end{aligned} \tag{A5}$$

Since $S(T_1+\tau)-S(T_1)=S'(T_1+\tau)-S'(T_1)$ holds for every $\tau$, both Eqs. (A6) and (A7) must also hold:

$$q/m = q'/m' \tag{A6}$$

$$(q-p)-2(q/m)Y(T_1) = (q'-p')-2(q'/m')Y'(T_1). \tag{A7}$$

From Eqs. (A6) and (A7), we get Eqs. (A8) and (A9).

$$(q'-p') = (q-p) + 2(q/m)Y'(T_1)-2(q/m)Y(T_1) \tag{A8}$$

$$Y'(T_1) = \frac{(q'-p')-(q-p)}{2(q/m)} + Y(T_1) \tag{A9}$$

Plug Eq. (A8) first and then Eq. (A9) into $S'(T_1)$, and after algebraic transformation, we get (A10).

$$\begin{aligned} S'(T_1) &= p'm' + (q-p)Y(T_1)-(q/m)[Y(T_1)]^2 \\ &\quad + \frac{(q'-p')^2-(q-p)^2}{4(q/m)}. \end{aligned} \tag{A10}$$

From $S'(T_1)=S(T_1)$ we get:

$$\begin{aligned} &p'm' + (q-p)Y(T_1)-(q/m)[Y(T_1)]^2 + \frac{(q'-p')^2-(q-p)^2}{4(q/m)} \\ &\quad = pm + (q-p)Y(T_1)-(q/m)[Y(T_1)]^2 \\ &\quad \Rightarrow p'm' + \frac{(q'-p')^2-(q-p)^2}{4(q/m)} = pm \\ &\quad \Rightarrow p'm'4(q'/m') + (q'-p')^2-(q-p)^2 = pm4(q/m) \\ &\quad \Rightarrow 4p'q' + (q'-p')^2 = 4pq + (q-p)^2 \\ &\quad \Rightarrow q'+p' = q+p \end{aligned} \tag{A11}$$

From Eq. (A11) and $S(T_1+\tau)=S'(T_1+\tau)$, we have the following:

$$\begin{aligned} &\frac{m(p+q)^2}{p}\frac{\exp[-(p+q)(T_1+\tau)]}{[(q/p)\exp[-(p+q)(T_1+\tau)]+1]^2} \\ &\quad = \frac{m'(p'+q')^2}{p'}\frac{\exp[-(p'+q')(T_1+\tau)]}{[(q'/p')\exp[-(p'+q')(T_1+\tau)]+1]^2} \\ &\quad \Rightarrow \frac{m}{p[(q/p)\exp[-(p+q)(T_1+\tau)]+1]^2} \\ &\quad = \frac{m'}{p'[(q'/p')\exp[-(p'+q')(T_1+\tau)]+1]^2} \\ &\quad \Rightarrow \frac{(q'/p')\exp[-(p+q)(T_1+\tau)]+1}{(q/p)\exp[-(p+q)(T_1+\tau)]+1} \\ &\quad = \left(\frac{m'p}{mp'}\right)^{1/2} \quad \text{is a constant.} \\ &\quad \Rightarrow (q'/p') = (q/p) \end{aligned} \tag{A12}$$

Eqs. (A6), (A11) and (A12) lead to the (A13).

$$m' = m, \quad p' = p, \quad q' = q. \tag{A13}$$

Eq. (A13) contradicts the original assumption that at least one of the parameters is different. Therefore $S(T)$ has to be the unique solution. The theorem is therefore proved. □

## Appendix B. Historical sources and data sources

*Historical sources:*

Bayus, B.L. (1992). Have diffusion rates been accelerating over time? *Marketing Letters* 3 (3), 215–226.

Bud, R., Niziol, S., Boon, T, & Nahum, A. (2000). *Inventing the modern world: technology since 1750*. New York: Dorling Kindersley.

Bunch, B. & Hellemans, A. (1993). *The timetables of technology: A chronology of the most important people and events in the history of technology*. New York: Simon and Schuster.

Desmond, K. (1987). *The Harwin chronology of inventions, innovations, discoveries: From pre-historic to present day.* London: Constable.

Dummer, G.W.A. (1983). *Electronic inventions and discoveries: Electronics from its earliest beginnings to the present*. New York: Pergamon Press.

DuVall, N. (1988). *Domestic technology: A chronology of developments*. Boston: GK Hall.

Fisher, D.E. & Fisher, M.J. (1996). *Tube: The invention of television*. Washington, D.C.: COUNTERPOINT.

Giscard D'Estaing, V. & Young, M (1993). *Inventions and discoveries: What's happened, what's coming, what's that?* Ann Arbor, MI: Books on Demand.

Giscard D'Estaing, V. (1985). *The world almanac book of inventions. Mahwah*. NJ: The World Almanac.

Kohli, R., Lehmann, D. R. & Pae, J. (1999). Extent and impact of incubation time in new product diffusion. *Journal of Product Innovation Management*, 16, 134–144.

Lifshey, E. (1973). *The housewares story: A history of the American housewares industry.* Chicago: National Housewares Manufacturers Association.

McNeil, I. (Ed.) (1990). *An encyclopedia of the history of technology.* London: Routledge.

Merchandising Week (1972). Special 50-year summary issue, February 28, 1972.

Smith, A. (Ed.) (1998). *Television: An international history.* New York: Oxford University Press.

Travers, B & Muhr, J. (Eds.) (1994). *World of invention*. Detroit: Gale Research.

Van Dulken, S. (2000). *Inventing the 20th century: 100 inventions that shaped the world*. New York: New York University Press.

Williams, I.T. (2000). *A history of invention: From stone axes to silicon chips*. New York: Check-mark Books.

Yu, J. (2002). *Factors affecting diffusion patterns for new products: A hierarchical Bayesian mixture model*. Ph.D. Dissertation, The University of Texas at Dallas.

*Data sources:*

Dealerscope

Dealerscope marketplace

Economic almanac

Electrical Merchandising

Electrical Merchandising Week

Electronic market data book (2003). Arlington, VA: Consumer Electronics Assn.

Historical statistics of the United States

In-stat

Kagan media services

Merchandising Week

Statistical abstract of the U.S.

Strategis

The information technology industry databook

Yankee group

## References


Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, *15*(5), 215−227.

Bass, F. M., Gordon, K., Ferguson, T. L., & Githens, M. L. (2001). DIRECTV: Forecasting diffusion of a new technology prior to product launch. *Interfaces*, *31*(3), S82−S93.

Bayus, B. L. (1992). The dynamic pricing of next generation consumer durables. *Marketing Science*, *11*(3), 251−265.

Bayus, B. L. (1993). High-definition television: assessing demand forecasts for a next generation consumer durable. *Management Science*, *39*(11), 1569−1585.

Bemmaor, A. C., & Lee, J. (2002). The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. *Marketing Science*, *21* (2), 209−220.

Dekimpe, M. G., Parker, P. M., & Sarvary, M. (1998). Staged estimation of international diffusion models: An application to global cellular telephone adoption. *Technological Forecasting and Social Change*, *57*, 105−132.

Dekimpe, M. G., Parker, P. M., & Sarvary, M. (2000). Multimarket and global diffusion. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New-product diffusion models* (pp. 49−73). Boston, MA: Kluwer.

Kohli, R., Lehmann, D. R., & Pae, J. (1999). Extent and impact of incubation time in new product diffusion. *Journal of Product Innovation Management*, *16*, 134−144.

Lawrence, K. D. (1981). New product forecasting diffusion modeling with scenario sensitivity analysis. In R. Leone (Ed.), *Proceedings, market measurement and analysis conference* (pp. 194−199). Austin, TX: University of Texas at Austin.

Lawrence, K. D., & Lawton, W. H. (1981). Application of diffusion models: Some empirical results. In Y. Wind, V. Mahajan, & R. N. Cardozo (Eds.), *New-product forecasting: Models and applications* (pp. 529−541). Lexington, MA: Lexington Books.

Lilien, G., Rangaswamy, A., & Van den Bulte, C. (2000). Diffusion models: Managerial applications and software. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New product diffusion models* (pp. 295−311). Boston, MA: Kluwer.

Lilien, G. L., Rao, A. G., & Kalish, S. (1981). Bayesian estimation and control of detailing effort in a repeat purchase diffusion environment. *Management Science*, *27*(5), 493−506.

Mahajan, V. (1975). Computers in hospitals: A diffusion study, Ph.D. Dissertation, The University of Texas.

Mahajan, V., Mason, C. H., & Srinivasan, V. (1986). An evaluation of estimation procedures for new product diffusion models. In V. Mahajan, & Y. Wind (Eds.), *Innovation diffusion models of new product acceptance* (pp. 203−232). Cambridge, MA: Ballinger Publishing.

Mahajan, V., Muller, E., & Srivastava, R. (1990). Determination of adopter categories by using innovation diffusion models. *Journal of Marketing Research*, *27*(1), 37−50.

Mahajan, V., & Peterson, R. A. (1985). *Models for innovation diffusion.* Beverly Hills, CA: Sage.

Mahajan, V., & Wind, Y. (1986). Innovation diffusion models of new product acceptance: A reexamination. In V. Mahajan, & Y. Wind (Eds.), *Innovation diffusion models of new product acceptance* (pp. 3−25). Cambridge, MA: Ballinger Publishing.

Putsis, W. P., & Srinivasan, V. (2000). Estimation techniques for micro diffusion models. In V. Mahajan, E. Muller, & Y. Wind (Eds.), *New product diffusion models* (pp. 263−291). Boston, MA: Kluwer.

Srinivasan, V., & Mason, C. H. (1986). Nonlinear least squares estimation of new-product diffusion models. *Marketing Science*, *5*(2), 169−178.

Sultan, F., Farley, J. U., & Lehmann, D. R. (1990). A meta-analysis of applications of diffusion models. *Journal of Marketing Research*, *27*, 70−77.

Van den Bulte, C. (2000). New product diffusion acceleration: Measurement and analysis. *Marketing Science*, *19*(4), 366−380.

Van den Bulte, C., & Lilien, G. L. (1997). Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Science*, *16* (4), 338−353.

Vanhonacker, W. R., Lehmann, D. R., & Sultan, F. (1990). Combining related and sparse data in linear regression models. *Journal of Business and Economic Statistics*, *8*, 327−335.

**Virtual Bass Model: Typos and Corrections**

| Page | Typo | Correction |
| --- | --- | --- |
| p 96 | $m^{'} - m = Y(\tau)$ | $m^{'} - m = -Y(\tau)$ |
| p 96 | $p_{Fs} = p - (q/m)Y(2T^{*}) = q$ (16) | $p_{Fs} = p + (q/m)Y(2T^{*}) = q$ (16) |
| p 98 | $U + p + q$ (V2) | $U = p + q$ (V2) |
| p 99 | In Table4, $q'$ for VBM: 63465935,<br>Difference for $q'$: –1.06617E-06,<br>Difference for $p'$: 6.61657E-08,<br>Difference for $m'$: –0.029426755. | In Table4, $q'$ for VBM: 0.63465935,<br>Difference for $q'$: 3.92E-04,<br>Difference for $p'$: 1.00E-09,<br>Difference for $m'$: 3.00E-02. |
| p 99 | The values for $p'$ differ only in 6 decimal places and for $q'$ differ only in 8 decimal places, … | The values for $p'$ differ from the 9th decimal place and for $q'$ differ from the 4th decimal place, … |
| p 105 | $S^{'}(T)$ by $I^{'}, p^{'}$, and $q^{'}$. | $S^{'}(T)$ by $m^{'}, p^{'}$, and $q^{'}$. |