



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
IMA Journal of Management Mathematics

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa43560>

Paper:

Goodwin, P., Dyussekeneva, K. & Meeran, S. (2013). The use of analogies in forecasting the annual sales of new electronics products. *IMA Journal of Management Mathematics*, 24(4), 407-422.
<http://dx.doi.org/10.1093/imaman/dpr025>

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

The use of analogies in forecasting the annual sales of new electronics products

Paul Goodwin*, Karima Dyussekeneva and Sheik Meeran

School of Management, University of Bath, Bath, BA2 7AY, United Kingdom.

November 2011

*Corresponding author: Email- mnspg@bath.ac.uk

Abstract

Mathematical models are often used to describe the sales and adoption patterns of products in the years following their launch and one of the most popular of these models is the Bass model. However, using this model to forecast sales time series for new products is problematical because there is no historic time series data with which to estimate the model's parameters. One possible solution is to fit the model to the sales time series of analogous products that have been launched in an earlier time period and to assume that the parameter values identified for the analogy are applicable to the new product. In this paper we investigate the effectiveness of this approach by applying four forecasting methods based on analogies (and variants of these methods) to the sales of consumer electronics products marketed in the USA. We found that all of the methods tended to lead to forecasts with high absolute percentage errors, which is consistent with other studies of new product sales forecasting. The use of the means of published parameter values for analogies led to higher errors than the parameters we estimated from our own data. When using this data averaging the parameter values of multiple analogies, rather than relying on a single most-similar, product led to improved accuracy. However, there was little to be gained by using more than 5 or 6 analogies.

Keywords: Bass model, diffusion models, new product forecasting, analogies.

1. Introduction

Forecasting the sales of products that have yet to be launched is an important problem for companies. In particular, forecasts of the future values of sales time series (e.g. sales in each of the first n years of a product's life) will guide decisions relating to future production capacity, marketing budgets, human resource planning and research and development. These forecasts can also be used to estimate the discounted future returns on the investment that will be needed to develop and market the new product. Surprisingly, given its importance, new product sales forecasting has received relatively little attention in the literature (Kahn, 2006).

One reason for this sparsity of research may be the difficulty of producing accurate period-by-period sales forecasts for new products. By definition, no time series data that is specific to the product will exist so that existing sales patterns cannot be extrapolated to estimate future sales. Moreover, in industries where the pace of technological development is rapid, product life-cycles may be shortened by the appearance of superior new products, but the timings of these events will themselves be difficult to predict. In addition, a distinction needs to be made between forecasting adoptions and forecasting sales. A consumer becomes an adopter of a new product as soon as they have purchased it once. In contrast, sales include both adoptions and additional purchases of the product. For example, consumers may buy several television sets for their home or buy a replacement for their original purchase when it wears out. Adoptions are usually easier to forecast than sales because the latter will also be dependent on the consumer's propensity to make multiple purchases and on estimates of how long a product will be retained before it is replaced. These factors can be taken into account by increasing the complexity of the forecasting model but this raises another issue. Managers may be sceptical of forecasts produced by complex models that they do not understand even if these forecasts can be shown to be reliable (e.g. see Taylor and Thomas, 1982). There is no point in producing an accurate, but complex, forecasting model if its output will be totally ignored by decision makers.

One widely recommended solution to the problem caused by the absence of past time series data for new products is to identify products which are similar to the new product that have been launched in the past (Thomas, 1985, Bass et al, 2001). The sales time series for these analogous products can then be used to identify a forecasting model that it is hoped will

accurately represent the future sales pattern of the product which is due to be launched. Of course, the merits of this approach depend crucially on the validity of the assumption that similar products, launched several years apart, will have similar sales patterns. They also assume that the relevant attributes that define similarity can be identified. For example, will similar sales patterns result from products that are similarly priced or those have similar functions or those which are launched under similar economic conditions, or will a combination of these attributes be needed to determine similarity?

In this paper we investigate the effectiveness of using analogies by examining their application in forecasting the annual sales of new consumer electronics products in the US market. Our forecasting model is the well known basic Bass model (1969). We have chosen this model because of its widespread use and also its relatively simplicity and transparent rationale which means that it is likely to be credible to managers.

2. The Bass Model

The basic Bass model is designed to reflect two key factors that determine whether a given consumer will purchase a new product in a given period: whether they are an ‘innovator’ or an ‘imitator’. Innovators tend to purchase new products relatively soon after they have been launched and are not influenced in their purchase decision by the behaviour of other consumers. Imitators, in contrast are influenced by the purchase decisions of others and their propensity to buy a new product will increase as the population of adopters grows. In some markets, there may be a second way in which a greater population of existing adopters has a stimulating effect on subsequent adoption. More adopters leads to more complementary products (e.g. apps for Smart phones) and more product support services thereby making the product more attractive. The basic Bass model is represented by the following differential equation.

$$\frac{dY_t}{dt} = p(m - Y_t) + q \frac{Y_t}{m} (m - Y_t) \quad (1)$$

where: m = the market saturation level.

Y_t = the cumulative number of adopters at time t

p = the coefficient of innovation

q = the coefficient of imitation

This model represents the adoption process in continuous time. Usually, the process is observed at discrete points in time. In this case, Bass suggests the following model.

$$Y_t = Y_{t-1} + p(m - Y_{t-1}) + q \frac{Y_{t-1}}{m} (m - Y_{t-1}) \quad (2)$$

A large number of extensions of this basic model have been proposed. For example, these extensions can taken into account marketing mix variables, like advertising expenditure (e.g. Simon and Sebastian, 1987) and replacement purchases (Islam and Meade, 2000). However, all of these enhancements increase the complexity of the model, which in turn is likely to increase the number of parameters that need to be estimated from data which may be sparse.

When a time series on adoptions exists the model can be fitted to the series using either ordinary least squares (OLS), as originally proposed by Bass (also see Franses, 2011), maximum likelihood estimation (MLE)(e.g. Schmittlein and Mahajan, 1982) or non-linear least squares (NLLS) (Srinivasan and Mason, 1986). Each of these methods has particular limitations. For example, Srinivasan and Mason (1986), found that that MLE seriously underestimated the standard errors of p , q , and m . NLLS requires a ‘good’ initial estimate of parameter values, otherwise it might converge on a local, rather than a global, optima. OLS was found by Schmittlein and Mahajan (1982) to give a poorer fitting model, when ‘best fit’ is defined in terms of mean absolute deviation and mean squared error, than models derived through MLE and NLLS.

3. Using analogies to estimate parameters for the Bass model

As indicated earlier, when no time series exists, because a product has yet to be launched, resort can be made to fitting the model to series for analogous products that have been launched in an earlier period. Most researchers recommend that this process should be used to estimate only p and q with m being determined by other means such as consumer intentions surveys, demographic data or management judgment (e.g. Tigert and Farivar, 1981, Bass et al, 2001). A few papers have suggested how the analogies might be selected to estimate p and q . Some have used management judgment. For example, Bass et al (2001) used a process that they termed ‘guessing by analogy’ to forecast subscriptions for satellite

television from 1994 to 1999. They asked managers to judge whether the pattern of subscriptions would be most similar to that of either Color TV sales in the 1960s or Cable TV subscriptions in the 1980s. The managers chose the latter. Guessing by analogy has also been used to forecast sales of a new drug (Lilien et al, 1981) and forecasts for high definition television (Bayus, 1993). However, the use of judgment to identify analogies may be subject to a number of deficiencies. For example, people may choose analogies simply because they are recent or easily recalled (Lee et al, 2007) or when they only have superficial similarity with the target (Holyoak and Thagard, 1995). Moreover, initially selected analogies may still dominate judgment even when more appropriate analogies are presented or where debiasing techniques are employed (Bolton, 2003).

An alternative to judgment is to view potential analogies in very broad categories (e.g. the industry to which the product belongs) and to use average published values of p and q for these categories. For example, Lilien et al (1999) provide estimated Bass parameters for US sales of 54 products belonging to five product groupings, such as agricultural equipment and electrical appliances. Lawrence and Lawton (1981) suggest that the parameter values should be based on two even broader categories- industrial products and consumer products -and they provide suggested values for these two product types.

Clearly, values estimated for such broad categories are unlikely to relate closely to the specific characteristics of a given product launch. Thomas (1985) therefore proposes a structured procedure for the selection and use of specific analogies. This involves identifying analogies by scoring them for similarity to the new product on a range of factors such as those relating to the economic situation (presumably at the time of the product' launch), the behaviour of buyers in relation to the product and the marketing strategy that will apply to the product. A set of the most similar products are then selected. Estimates of p and q values for the target product are then derived by taking a weighted average of the p and q values for the selected analogies. The weights reflect consumers' utilities for the analogies, based on the extent to which they have attributes in common with the target product. Thomas only applied his proposed method in a very limited form but his approach appears to raise a number of problems. First the factors used to identify the initial set of analogies are only vaguely defined (e.g. "buying situation", "segmentation etc") and their usefulness does not appear to have been tested in a forecasting context. Second it requires estimates of consumers' preferences. For this he suggests using a primary market study or a combination of

secondary data and expert judgment. The first source would not only be expensive, but it would also be subject to a number of survey-related biases. The second source would potentially suffer from a range of judgmental biases many of which may apply even when the assessor is a recognised expert.

In contrast to Thomas's use of expert judgment or consumer surveys, Ilonen et al (2006) present an automatic procedure for identifying analogies based on a self-organizing map (or Kohonen map). This was used to estimate cell phone usage in different countries. The map organized countries according to their similarities on multiple dimensions and estimates for a given country were obtained by combining Bass models for similar countries. This approach can be regarded as a form of nearest neighbour analysis (Nikolopoulos et al, 2007). Nearest neighbour analysis requires an assessment to be made of the number of neighbours that should be identified as suitable analogies. If more neighbours are used then this should filter out the noise associated with the data from individual analogies, but it will probably also mean that less similar analogies will be included in the set which is used to produce the forecasts.

An alternative is to apply regression analysis to the data that is available on the analogies to determine how p and q are related to the characteristics of the different products. The appropriate values of p and q for the new product can then be estimated by using equations of the form shown below.

$$\hat{p} = \hat{\alpha} + \sum_{i=1}^n \hat{\beta}_i x_i \quad (3)$$

$$\hat{q} = \hat{\phi} + \sum_{i=1}^m \hat{\theta}_i x_i \quad (4)$$

where: \hat{p} = the estimate of p that will be used to produce forecasts for the new product

\hat{q} = the estimate of q

x_i = the value of characteristic i for the new product

α, β, ϕ and θ are the population regression parameters

n and m = the number of characteristics that are thought to determine p and q, respectively.

Approaches similar to this have been used by Srivastava et al (1985) to forecasting consumers' adoption of investments and by Gatignon et al (1989) to model diffusion of innovations across different countries. However, Ilonen et al (2006) and Nikolopolous et al (2007) found that nearest neighbour analysis outperformed linear regression analysis in forecasting. This may be because nearest neighbour analysis can handle situations where the similarity of an analogy to a target is dependent on a complex interaction of the analogy's attributes rather than a simple linear combination.

Even if a set of possible analogies can be identified, two further problems can arise in many new product forecasting situations. These occur when the time series data for the analogy either does not include observations starting at the product's launch date (left-hand data-truncation) or excludes observations from the later stages of the product's life, including the period when adoption reaches its peak (right-hand truncation). Both data omissions can lead to bias in the estimates of p and q . Left-hand truncation can lead to substantial under estimates of the time that it will take a product to reach its peak sales (Jiang et al, 2006). Jiang et al present a method for mitigating the bias and a table of modified p and q estimates for 39 products that can be used for 'guessing by analogy'. Right-hand data truncation tends to lead to under-estimates of p and overestimation of q when estimation is based on non-linear least squares (Van den Bulte and Lilien, 1997, Bemmaor and Lee, 2002). For analogy-based forecasting this type of truncation is likely to be particularly problematical in dynamic markets. In these cases a choice may need to be made between an analogous product that has been launched relatively recently and an older product. The former may closely reflect current consumer behaviour but it will have a relatively short sales history; the latter will have a longer history but may not be relevant to current market conditions.

One common feature of much of the research reported above is the absence of the use of extensive out-of-sample data sets to test forecast accuracy. This may reflect the general difficulties of obtaining long time series data relating to products following their launch or the complete absence of such a data set if the study has been carried out close to the launch date. For example, Bayus (1993) had no out-of-sample data available at all and assessed the reliability of forecasts by assessing their consistency with other published forecasts that were available at the time. Thomas (1985) tested forecasting accuracy on only two data points, while Bass et al (2001) reported promising results based on only 5 out-of-sample observations. Ilonen et al (2006) did not produce true forecasts: the time series for the

analogous markets covered the same periods as the target markets –they argued that their results indicated ‘maximal approachable accuracy’. All of this is a concern because, as Tashman (2000) has pointed out, extensive out-of-sample testing is needed to allow reliable inferences to be drawn about the accuracy of a given method.

In the analysis that we describe next we use observations from a large sales data base to compare the accuracy of four different methods of using analogies (and several variants of these methods) to produce time series forecasts for newly launched products.

4. The data set and accuracy measurement

We obtained the *FastFacts Historical Sales Data* database produced by the US Consumer Electronic Association (CEA). This contained sales data of 97 electronic goods, such as TVs, radios, CD players and cellular phones that were launched commercially on the US market during the period 1946 and 2007. We categorised products that were launched before 1995 as potential analogies (with only sales up to 1995 being used to estimate parameters) and products that were launched after 1995 as targets. Products were not used in the analysis when sales were recorded as zero for all years. Also products were rejected as analogies when: i) there were less than 5 sales figures available (Heeler and Hustad (1980) have suggested that the Bass model is not applicable when shorter data sets are available), ii) it was evident that the launch date preceded the starting date of the available data by several years (e.g. data on turntables was only available from 1980), iii) a Bass model with sensible parameter values could not be fitted to the sales time series (e.g. negative parameter values were obtained in some cases, as a result of unusual sales patterns or outlying observations). This is reasonable as a manager would be able to screen out such unusual potential analogies before making forecasts. This process yielded 23 potential analogies and 21 target products. Table 1 gives details of the start dates of the time series for targets and analogies. Also shown are the number of observations (n) that were used to fit the Bass models for the analogies and to assess the accuracy of the forecasting methods for the targets. In total 210 observations were used to assess accuracy.

| Potential Analogous products | Date of first sales | | Target Products | Date of first sales | |
|--------------------------------------|---------------------|----|---------------------------------------|---------------------|----|
| | figure | n | | figure | n |
| Analog Color TV with stereo | 1984 | 11 | Cable/Multi-System Operator Receivers | 2003 | 7 |
| Analog Handheld LCD Color TV | 1985 | 10 | Caller ID Devices | 1996 | 14 |
| Analog TV/VCR Combinations | 1990 | 5 | Digital Cameras | 1996 | 14 |
| Analogue TV | 1954 | 41 | Digital Direct-View Sets & Monitors | 2000 | 10 |
| Blank audio cassettes | 1985 | 10 | Digital Front Projection TV | 2002 | 8 |
| Blank video cassettes | 1982 | 13 | Digital projection sets | 2004 | 11 |
| Camcorders | 1985 | 10 | Digital TV Sets & Displays | 1998 | 12 |
| Cellular Phones | 1984 | 11 | Digital Video Recorders (DVRs) | 2003 | 7 |
| Corded Telephones | 1982 | 13 | DVD Players/Recorders | 1997 | 13 |
| Cordless Telephones | 1980 | 15 | Family Radio Devices | 1997 | 13 |
| Fax Machines | 1987 | 8 | HDTV | 2003 | 7 |
| Home Radios | 1950 | 45 | Home Theater-in-a-Box | 1996 | 14 |
| Laserdisc Player | 1985 | 10 | Personal Digital Assistants | 2000 | 10 |
| Monochrome TV | 1946 | 49 | Personal Video Recorders (PVR) | 1999 | 9 |
| Personal Wordprocessors | 1982 | 13 | Plasma DTV | 1999 | 11 |
| Portable CD Equipment | 1987 | 8 | Portable MP3 Players | 1999 | 11 |
| Portable Headset Audio | 1985 | 10 | Portable and Transportable Navigation | 2000 | 10 |
| Portable Tape and Radio/Tape Players | 1983 | 12 | Set-top Internet Access Devices | 1997 | 9 |
| Rack Audio Systems | 1980 | 15 | Smartphones | 2003 | 7 |
| Telephone Answering Devices | 1982 | 13 | TV/DVD Combinations | 2002 | 6 |
| Total CD Players | 1983 | 12 | VOIP Adapters | 2003 | 7 |
| VCR Decks | 1974 | 21 | | | |
| Videocassette Players | 1985 | 10 | | | |

TABLE 1. Details of analogies and target products

A large number of measures are available to assess the accuracy of sales forecasts. Measures like the mean squared error are inappropriate when accuracy is being measured across series where sales have different scales (e.g. hundreds of units for some products and millions for others). Also managers often prefer measures based on percentages. However, the widely used mean absolute percentage error (MAPE) is distorted when some periods have very low sales (e.g. see Goodwin and Lawton, 1999). For new products this distortion is likely to occur because very low sales are typically observed in the early stages of a product's life. To counter this we measured accuracy using a modified version of the MAPE (the MMAPE) which assessed absolute forecast errors relative to the mean sales of the product during its observed life, rather than individual period-by-period sales. The formula for the MMAPE is

$$\text{MMAPE} = \frac{100 \sum_{t=1}^n \frac{|Y_t - F_t|}{A}}{n} \quad (5)$$

where: Y_t = the actual sales for period t

F_t = the forecast sales for period t

A = the mean sales over the product's observed life

n = the number of periods in the product's observed life.

A Bass model (2) was fitted to the time series for each analogy using non-linear least squares to obtain estimates of p and q using only data up to 1995 (starting estimates were determined by OLS to reduce the danger of non-global optima being identified). However, it would be unreasonable to use the value of m estimated for an analogy as an estimate of the market saturation level for a target. For example, an analogy and a target may have similar diffusion patterns, but very different levels of saturation. We therefore assumed that the value of m for the targets could be determined accurately through methods like consumer intentions survey or demographic analysis. To obtain proxies for these values we fitted Bass curves to the time series for the target products and assumed that the saturation level would be equivalent to the resulting m estimates. Clearly, because we were using data that would not be known at the time of the forecast to estimate m , the accuracy that we report for our forecasts may be higher than will be the case in an actual application of our methods. For example, m estimates based on consumer surveys are unlikely to be perfectly accurate and in some cases may have significant errors (Morwitz et al, 2007). However, this potential bias applied equally to all of the methods we compared.

5. Forecasting methods

We compared four main methods of using analogies to produce time series forecasts for the target products using a Bass model. For each method the p and q values were determined as indicated below.

1. *Published values.* Mean p and q values for 13 consumer electronic products published by Lilien et al (1981) in their Exhibit 1a. Note that six of the sales series used to estimate the parameters contained an observation for 1996 which would not have been known at the time

of the launch of 3 of our target products. This may cause the accuracy of this approach to be slightly over estimated. We also used mean p and q values for nine consumer electronics products published by Jiang et al (2006) –recall that these values were modified to remove the bias arising through left-hand truncation bias. In this case, estimates for just two products were partly based on post 1995 data so these were excluded from the calculation of the mean.

2. *Random selection of analogies.* We randomly selected k analogies from those listed in table 1 and used their mean p and q values to produce the forecast. This procedure was repeated 1000 times and the mean MMAPES recorded. This approach was investigated for each k value from 1 to 22. For k= 23 all of the analogies were selected so the random selection was not required. The standard errors of the estimates of the mean MMAPES were all below 0.53 so 1000 repetitions was judged to be sufficient.

3. *Nearest neighbour analysis.* The similarity of a given analogy to a target product was based on the attributes listed below. These attributes were identified through other studies and a formal brainstorming session carried out by the three researchers.

- a. Whether or not there would have been a threat of a substitute product at the year of launch
- b. Whether or not the product was portable
- c. Whether or not the product was highly useful/compelling, so it could not be substituted and was unique in the sense of practical application. For example, a car satellite navigation system
- d. Whether or not the average time before the product was replaced with a new version was likely to be less than or equal to 5 years (Replace)
- e. Whether or not the major function of the product was to record still and moving pictures
- f. Whether or not the primary use of the product was to facilitate live two-way communication between at least two parties
- g. Whether or not the primary function of the product was to allow the user to both to record and playback music
- h. Whether or not the primary use of the product was to supply entertainment produced by a party other than the user (Entertain)
- i. Whether or not the product would be useful to a typical small business

- j. The estimated number of days that a person on average income would have to work to buy the product in its year of launch (Days)
- k. The logarithm of the number of observations available on the product's sales. (log(n))
- l. The estimated market saturation level (m)
- m. The product's date of launch

For characteristics (a) to (i) the researchers first independently assessed both the analogies and the products for the presence or absence of that characteristic. These were represented by dummy variables in the analysis. Any differences between these assessments were subsequently resolved through discussion at a meeting. Data on attribute (j) was obtained from published sources. Attribute (k) was included to take into account right-hand truncation bias.

The use of both 1 (NN1), 3 (NN3) and 5 (NN5) nearest neighbours was investigated. In the case of 3 nearest neighbours a weighted mean of the p and q values of the nearest neighbours was taken with the weights of 0.5, 0.25 and 0.25 being applied to the neighbours based on their closeness to the target. In the case of 5 nearest neighbours a simple mean of the p and q values was used. These three applications replicated the nearest neighbour approaches used by Nikolopoulos et al (2007). However, we used Gower's dissimilarity coefficient (Gower, 1971) to measure the distance between a given analogy and target, because it enables similarity to be assessed when attributes are measured on different scales (e.g. binary, continuous and categorical).

4. *Regression analysis* of the form shown in (2) and (3) was applied to the analogous products using OLS to estimate the relationship between p and q and attributes (a) to (l). Stepwise regression was used to identify which attributes should be included in the regression models. The model fit was improved when the logs of p and q were used as the dependent variables and in some cases the logs of independent variables also improved the fit. The resulting models were:

$$\text{Log}(p) = - 3.71 - 0.000002 m - 0.0400 \text{ Days} \quad (6)$$

(R² = 62.7%, all coefficients significant at p < 0.015)

$$\begin{aligned} \text{Log}(q) = & - 0.023 - 1.13 \log(n) + 0.179 \log(\text{Days}) + 0.574 \text{Entertain} \\ & + 0.102 \log(m) - 0.341 \text{Replace} \quad (7) \\ (R^2 = & 84.3\%, \text{ all coefficients, except the constant, significant at } p < 0.03) \end{aligned}$$

We compared the accuracy of these four methods with two benchmarks.

Benchmark 1: *Perfect information on sales*. The optimum models was fitted to the available sales series for each target product. The resulting MMAPE indicates the extent to which forecasts errors arise because the underlying sales pattern failed to conform to a Bass model or because the sales pattern was subject to noise.

Benchmark 2: *Perfect information on the best analogy*. Here we used the p and q values of the analogy that led to the most accurate forecasts for each target product. No analogy is likely to have a sales pattern that perfectly replicates that of a target product and the MMAPE here indicates the extent to which errors arise because of differences between the two products. MMAPEs that are higher than this benchmark will reflect the extent to which errors arise because an inferior analogy has been chosen.

6. Results

Table 2 shows the MMAPEs for the target products for the different forecasting strategies and the two benchmarks. The means in the last two rows are weighted to take into account the number of forecasts that are being made for each product. The benchmarks show that, even with perfect information on sales, a Bass model yields a mean MMAPE of 13.7% suggesting that significant component of forecasting error will be the inability of the Bass model to represent sales patterns perfectly. This may be partly arise because the model is primarily intended to represent adoptions, rather sales. The second benchmark shows that, even if we could always identify the analogy in the data base which will yield the lowest MMAPE for each target product, then the mean MMAPE would still be 30.1% which indicates the extent of the variation between the sales patterns of the different products.

The results for the forecasting methods in table 2 suggest that Personal Video Recorders (PVR) is an outlier with several MMAPEs being well over 200%. When this product was removed a Freidman's test indicated that there was a significant difference between the

accuracy of the eight forecasting methods displayed in table 2 ($p < 0.00001$). An examination of the table suggests that the methods fall into three broad categories. The worst overall performance is obtained by using the means of published p and q values for consumer electronics. The use of a single analogy from the database provides the next level of accuracy. However, surprisingly, there is no significant difference between random selection of this analogy and its identification through nearest neighbour analysis (Wilcoxon test, $p=0.33$). The methods which are based on more than one analogy offer the highest level of accuracy, but there appears to be little to choose between them. All have MMAPEs of around 50%. A Friedman test did indicate that using all analogies yielded less accurate forecasts than NN3, NN5 and regression analysis ($p=0.041$) but this should be interpreted with caution given that multiple comparisons are being made.

If multiple analogies are preferable, how many should be selected? Figure 1 displays the MMAPES that were obtained when k analogies were randomly selected from those listed in table 1 and the forecasts were based on their mean p and q values. The MMAPEs displayed are means based on repeating the random selection process 1000 times. It can be seen that the expected MMAPEs falls as the number of analogies selected increases but the greatest benefits are obtained by increasing k from 1 up to 5 or 6. The standard deviations (SDs) of the results derived from the simulation are also shown. These can be used to assess the risk of basing forecasts on a single random selection. Again little is to be gained by selecting more than 5 or 6 analogies

7. Discussion

Three main conclusions follow from the above results. First, forecasting future time series values for new products is a challenging task and, when using analogies, it is difficult to obtain MMAPEs below about 50%. This level of accuracy is consistent with forecast performance that has been reported in companies. For example, Tull (1967) studied new product forecasting in 16 firms and found that the MAPE for new consumer products was 49%. However, the results appears to be better than the mean forecasting accuracy for 'new to the world' products reported by Kahn (2002). The latter paper reported an 'average percentage accuracy' achieved of 40.36% which we presume would translate roughly into MMAPE of about 60%. The second conclusion is that using the means of published

parameter values for particular product groups was not conducive to accurate forecasting. Finally, using means of the parameter values of several analogies, based on our own data, was preferable to using a single analogy, though the choice of how these multiple analogies were selected did not appear to be important. We now explore these findings in more depth.

Why did the use of analogies lead to forecasts with such high absolute percentage errors? One reason is the heterogeneity of the sales patterns of the products. Figure 2 shows the estimated p and q values for the analogies and the target products (indicated by T). This shows that the q values, in particular, have a considerable range. Moreover, some target products have p and q values that are very different from any analogy. It also appears that the q values of the targets tend to be higher than those of the analogies. Two reasons may account for this phenomenon. First, there may have been a real change in q over the years, perhaps reflecting the increased power of 'the word of mouth' as a result of the internet. Other researchers have also reported increases in q over the years, but Decker and Gribba-Yukawa (2009) suggest that this may be because modern high-technology products experience faster market growth than their predecessors because of rapidly dropping prices. Whatever the cause, such changes would clearly limit the effectiveness of using the q values of pre-1996 products as analogies for those launched on or after this date. A second possibility is that the apparent change is simply a result of the number of observations used in the estimation. As discussed earlier, Van den Bulte and Lilien (2007) have shown that q tends to be overestimated when relatively few observations are available and non-linear least squares is being used to fit the model. In the case of q this upward bias tends to decrease systematically as more observations become available.

| Target product | Perfect information on sales | Use of best analogy | Mean values from Lilien et al | Mean values from Jiang et al | Random selection of one analogy | Use of all analogies | NN1 |
|---------------------------------------|-------------------------------------|----------------------------|--------------------------------------|-------------------------------------|--|-----------------------------|------------|
| Cable/Multi-System Operator Receivers | 3.1 | 8.1 | 215.2 | 72.4 | 59.5 | 11.7 | 72.7 |
| Caller ID Devices | 8.0 | 21.1 | 121.4 | 54.8 | 56.6 | 29.5 | 31.7 |
| Digital Cameras | 6.8 | 22.8 | 140.4 | 31.3 | 66.6 | 42.6 | 60.3 |
| Digital Direct-View Sets & Monitors | 36.7 | 51.5 | 68.3 | 85.0 | 71.8 | 64.5 | 66.9 |
| Digital Front Projection TV | 7.9 | 11.8 | 64.0 | 81.3 | 53.6 | 38.9 | 44.4 |
| Digital projection sets | 13.4 | 42.6 | 39.8 | 81.8 | 69.0 | 63.0 | 64.6 |
| Digital TV Sets & Displays | 7.5 | 44.7 | 143.1 | 56.0 | 74.1 | 56.0 | 57.2 |
| Digital Video Recorders (DVRs) | 17.2 | 39.1 | 70.1 | 87.9 | 64.4 | 54.9 | 69.5 |
| DVD Players/Recorders | 15.3 | 15.5 | 79.3 | 54.5 | 50.5 | 26.7 | 113.0 |
| Family Radio Devices | 17.0 | 20.2 | 54.7 | 66.3 | 52.6 | 35.7 | 20.2 |
| HDTV | 3.0 | 22.5 | 47.7 | 93.2 | 77.5 | 76.0 | 78.0 |
| Home Theater-in-a-Box | 15.3 | 15.1 | 104.8 | 44.6 | 48.4 | 14.4 | 15.1 |
| Personal Digital Assistants | 13.7 | 24.6 | 37.4 | 77.5 | 57.8 | 45.0 | 38.6 |
| Personal Video Recorders (PVR) | 5.3 | 35.8 | 401.5 | 42.7 | 140.0 | 115.5 | 164.0 |
| Plasma DTV | 14.6 | 30.9 | 115.9 | 73.5 | 74.7 | 62.3 | 74.2 |
| Portable MP3 Players | 14.8 | 34.8 | 114.2 | 73.2 | 76.8 | 67.0 | 73.4 |
| Portable and Transportable Navigation | 40.8 | 65.1 | 129.3 | 82.2 | 88.9 | 81.8 | 83.4 |
| Set-top Internet Access Devices | 9.6 | 45.4 | 41.4 | 90.6 | 80.7 | 78.6 | 76.3 |
| Smartphones | 2.1 | 45.1 | 64.0 | 88.7 | 67.9 | 59.5 | 67.0 |
| TV/DVD Combinations | 16.6 | 25.8 | 28.0 | 94.9 | 79.2 | 78.5 | 80.7 |
| VOIP Adapters | 11.6 | 15.1 | 42.3 | 92.5 | 76.5 | 74.9 | 69.1 |
| Weighted means | 13.7 | 30.1 | 103.1 | 69.0 | 69.3 | 53.3 | 64.9 |
| Weighted Means excluding PVR | 14.0 | 29.8 | 89.7 | 70.2 | 66.1 | 50.5 | 60.5 |

TABLE 2. MMAPEs for the target products

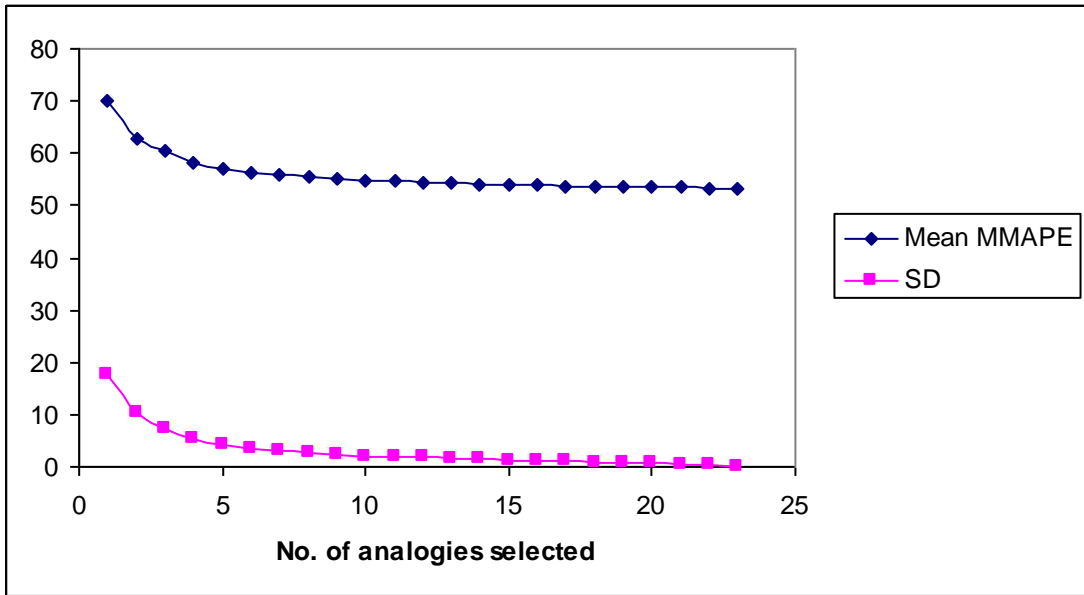


FIG. 1. Random selection of analogies and MMAPEs

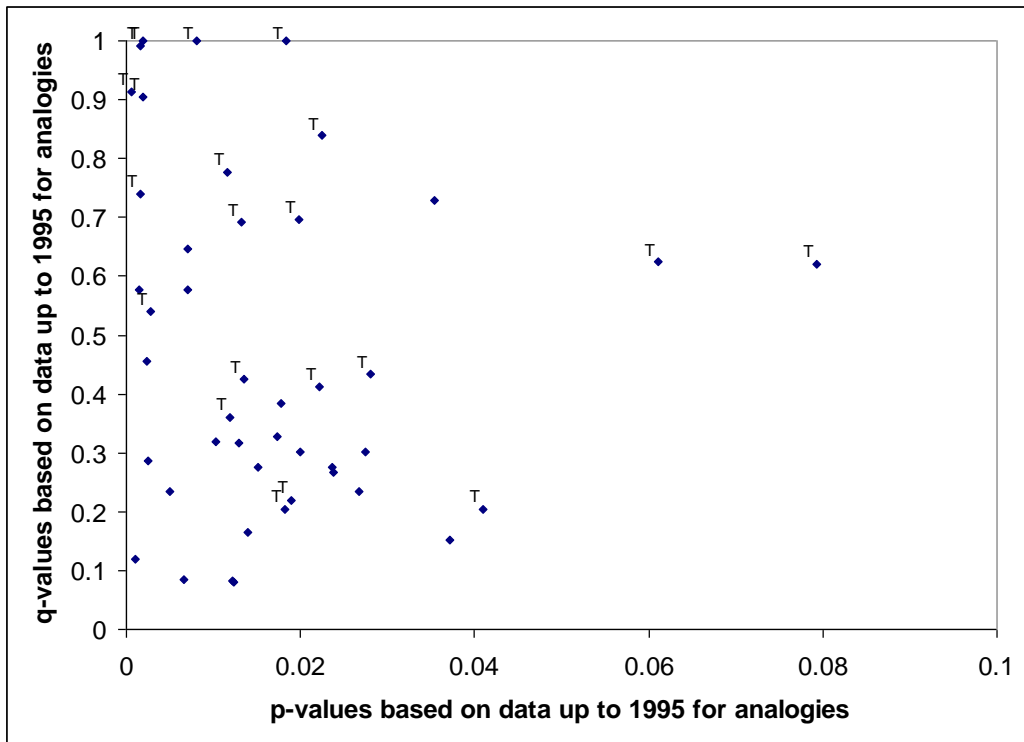


FIG. 2. Estimated p and q values for analogies and target products

To test these two possible explanations we fitted the following regression models to the data:

$$\hat{p} = b_0 + b_1D + b_2n + b_3 D n \quad (8)$$

$$\hat{q} = b_3 + b_4D + b_5n + b_6 D n \quad (9)$$

where: $D = 0$ if the product launch date was before 1996 and 1 otherwise,

n = the number of observations used to estimate p and q ,

and the b_i are regression coefficients.

Here a significant value for b_1 or b_4 would suggest an increase in the value of p or q , between the pre-1996 and post 1995 products irrespective of the number of observations used to estimate these parameters. A significant value for b_2 or b_5 would indicate that the estimate is influenced by the number of observations. Finally, a significant value for b_3 or b_6 would indicate that the effect of the number of observations on the estimate has changed between the pre and post 1995 products.

The model obtained for \hat{p} was:

$$\hat{p} = 0.0201 + 0.0271 D - 0.000335 n - 0.00249 D n \quad (10)$$

(0.000) (0.062) (0.216) (0.062)

$$R^2 = 15.6\% \quad F_{3,40} = 2.46 \quad (\text{significant at } p = 0.076)$$

where the values in parentheses are the levels of significance for the regression coefficients.

It can be seen that there is only weak evidence (at the 10% levels of significance) of any of the above effects.

For \hat{q} the following model was obtained.

$$\hat{q} = 0.449 + 0.485 D - 0.00862 n - 0.0201 D n \quad (11)$$

(0.000) (0.022) (0.031) (0.29)

$$R^2 = 45.8\% \quad F_{3,40} = 11.25 \quad (\text{significant at } p < 0.001)$$

This clearly suggests that there has been a mean increase in q between the two periods of 0.485, which is unrelated to the number of observations used in the estimation. Also, consistent with the findings of van den Bulte and Lilien (2007), an increase in the number of observations is associated with a reduction in the estimate of q . However, there is no evidence that this biasing effect, itself, has changed between the two periods.

Both the nearest neighbour analysis and the regression analysis included the number of observations used to estimate p and q so these methods should have taken this potential bias into account. However, neither method could take into account the underlying increase in q , which would not have been known in 1995. As a result the regression-based forecasts of q underestimated the observed values by a mean amount of 0.26.

Why did the use of the means of p and q values estimated in the two published studies for consumer electronic products lead to such inaccurate forecasts? These p and q estimates for individual products had the advantage of being based on long time series and, in the case of Jiang et al (2006), biases arising from left-hand truncation bias were removed. However, because they were based on series that up to the mid 1990s their q values tended to be too low (e.g. the mean level of under estimation was 0.19 for the Jiang et al data). In addition, the published results were presented by the researchers with the intention that forecasters would select an appropriate individual analogy from the tables, rather than a mean of the parameter estimates for an entire product group (though of course our results have suggested that this selection is likely to be problematical). Our results may therefore suggest that the mean p and q values for a product group may not provide reliable estimates of the p and q values for a specific product within that group.

To investigate this we used Lilien et al's (1999) data which provides estimated Bass parameters for US sales of 54 products belong to five product groupings (Agricultural equipment, Medical equipment, Production technology, Electrical appliances and Consumer electronics). The launch dates of the products range from 1815 to 1991 and the number of years of observed sales range from 6 to 151 with the most recent observed sales year being 1996. For this data a general linear model was used to test for significant effects on p and q estimates of (i) product launch date, (ii) the market saturation level (m) iii) the log of the number of observations and iv) the product group. The results revealed no significant difference in the p and q values between the product groups –despite the wide diversity of

these groups. Thus within-product group variation in p and q may be as large as between-group variation, after the other variables have been taken into account. This suggests that analogies may need to be identified at a deeper level than simply membership of the same product group.

However, this does not explain why the mean p and q values of all the analogies in our data set of consumer electronics products yielded MMAPEs that were closer to those of NN3, NN5 and regression analysis and which were better than the means of the published values. Although our data set included 23 analogous products, whereas the means from the published studies were based only 9 and 13 products this seems, this does not seem to account for the different levels of accuracy. As shown above only small gains in accuracy were achieved when more than 5 or 6 analogies were used. We can only speculate that the causes of the relatively poor performance of the means of published parameters may be more subtle. For example, perhaps all the data in a single database is recorded according to the same rules and definitions. Or perhaps the relative recency of our series was an advantage. The mean start dates of the published series were 1964 and 1974 while our analogies had a mean start date of 1979. Also the use of secondary data denies one the opportunity to filter out products that may display unusual sales patterns.

We reported earlier that NN1 did not yield significantly more accurate results than using a single randomly selected analogy. There was also no strong evidence that NN3 and NN5 outperformed the random selection of 3 and 5 analogies respectively. However, the regression analysis indicated that at least some of the factors used in the nearest neighbour analysis were significantly related to p and q . It may have been that other, irrelevant, factors cancelled out the benefits of including these factors. Unlike regression, the nearest neighbour approach, as we applied it, did not include a method for filtering out irrelevant factors. Nor did it allow an assessment to be made of whether different weights should be assigned to the factors. However, while regression did not suffer from these limitations, it assumed a linear relationship between the parameter values and the factors so that any interactions between them or other non-linearities could not be taken into account.

8. Conclusions

Our analysis has a number of limitations. We used secondary data on the annual national sales of products, rather than data on specific brands or generations of a product and our data did not allow us to distinguish between adoptions and sales. Also our data did not contain information on the marketing strategies that accompanied the launch of new products or on barriers to entry or regulatory conditions. Further, our analysis was confined to consumer electronic products, though as we saw earlier, for durable products, variations of sales patterns between product groups may be no greater than variation within groups. In addition, we used proxy estimates of the market saturation level which were derived from information that could not have been known at the time when the forecast was made (though this advantage was common to all of the forecasting methods that were being compared). For many products data on all the years up to market saturation was either not available (because the products had been launched relatively recently) or could not be used (because only the sales figures for the analogies that pre-dated the launch of the targets could be used). However, this is likely to be a problem for analogy-based new product forecasting in general where the need to use recent analogies means that their full sales history will not yet have evolved.

We also rejected products as potential analogies where less than five years of sales observations were available because forecasts from a Bass model are likely to be unreliable under these conditions. However, this can lead to a survivor bias, where forecasts of the sales of new products are based only on the sales patterns of successful products or products that have a long life-cycle. Our analysis is therefore not applicable to situations where much shorter product life cycles are common (e.g. see Wu et al, 2006).

Given these caveats, our analysis has yielded a number of practical conclusions. First, forecasters and researchers often routinely suggest the use of analogies when the Bass model is to be applied to new product forecasting. This analysis has shown that the identification and use of analogies is neither trivial nor easy and that it is by no means guaranteed to produce reliable forecasts. In a dynamic world analogies that were launched under different economic or market conditions may provide little useful information for the assessment of sales patterns of new products. Second, using means of parameter values in published studies is risky and may lead to high forecast errors. Third, using the parameter values of a single

analogy, however, it was identified is also likely to lead to less accuracy than the mean of several analogies, even if these are randomly selected from the product group. However, it appears that there is little to be gained in average accuracy by using more than about 5 or 6 analogies. Using regression analysis to forecast p and q values offers an alternative approach, if there are sufficient analogies available, but there may be a need for the model to incorporate non-linear relationships between the factors and the parameter values. Finally, given the apparent increases in q over time, either relatively recent analogies should be used (despite the relatively short histories associated with them) or, if older analogies are used, the possibility of adapting q estimates so that they incorporate these increases should be explored.

References

- BASS, F.M. (1969) A new product growth model for consumer durables. *Management Science*, **15**, 215-227.
- BASS, F.M., GORDON, K., FERGUSON, T.L. & GITHENS, M.L. (2001) DIRECT TV: Forecasting diffusion of a new technology prior to product launch. *Interfaces*, **31**, S92-S93.
- BAYUS, B.L. (1993) High-definition television: assessing demand forecasts for a next generation consumer durable. *Management Science*, **39**, 1569-1585.
- BEMMAOR, A.C. & LEE, J. (2002) The impact of heterogeneity and ill-conditioning on diffusion model parameter estimates. *Marketing Science*, **21**, 209-220.
- BOLTON, L.E. (2003) Stickier priors: the effects of nonanalytic versus analytic thinking in new product forecasting. *Journal of Marketing Research*, **XL**, 65-79.
- DECKER, R. & GNIBBA-YAKAWA, K. (2010) Sales Forecasting in High-Technology Markets: A Utility-Based Approach. *Journal of Product Innovation Management*, **27**, 115-29.
- FRANSES, P.H. (2011) Marketing and sales. *The Oxford Handbook of Economic Forecasting* (M.P. Clements & D.F. Hendry eds). Oxford: Oxford University Press, pp 673-689.
- GATIGNON, H., ELIASHBERG, J. & ROBERTSON, T.S. (1989) Modeling multinational diffusion patterns: an efficient methodology. *Marketing Science*, **8**, 231-247.
- GOODWIN, P. & LAWTON, R., (1999) On the asymmetry of the symmetric MAPE. *International Journal of Forecasting*, **15**, 405-408.
- GOWER, J. C. (1971) A general coefficient of similarity and some of its properties. *Biometrics*, **27**, 623-637.
- HEELER, R. & HUSTAD, T. (1980) Problems in predicting new product growth for consumer durables. *Management Science*, **26**, 1007-1020.
- HOLYOAK, K.J. & THAGARD, P. (1995) *Mental Leaps: Analogy and Creative Thought*. Cambridge: MA: MIT Press, pp. 101-137.
- ILONEN, J., KAMARAINEN, J-K, PUUMALAINEN, K., SUNDQVIST, S. & KÄLVIÄINEN, H. (2006) Toward automatic forecasts for diffusion of innovations. *Technological Forecasting and Social Change*, **73**, 182-198.
- ISLAM, T. & MEADE, N. (2000) Modelling diffusion and replacement. *European Journal of Operational Research*, **125**, 551-570.
- JIANG, Z., BASS, F.M. & BASS, P.I. (2006) Virtual Bass model and the left-hand data-truncation bias in the diffusion of innovation studies. *International Journal of Research in Marketing*, **23**, 93-106.

KAHN, K.B. (2002) An exploratory investigation of new product forecasting practices. *The Journal of Product Innovation Management*, **19**, 133 – 143

KAHN, K.B. (2006) *New Product Forecasting: An Applied Approach*, Armonk, NY: ME Sharpe.

LAWRENCE, K.D. & LAWTON, W.H. (1981) Applications of diffusion models: some empirical results. *New Product Forecasting* (Y. Wind, V. Mahajan, & R Cardozo eds). Lexington, MA: Lexington Books, pp529-541.

LEE, W.Y., GOODWIN, P., FILDES, R., NIKOLOPOULOS, K. & LAWRENCE, M. (2007) Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, **23**, 377-390.

LILIEN, G., RANGASWAMY, A. & VAN DEN BULTE, C. (1999) Diffusion Models: Managerial Applications and Software. Institute for the Study of Business Markets, Pennsylvania State University, Report 7-1999.

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**, 493-506.

MORWITZ, V.G., STECKEL, J.H., GUPTA, A. (2007) When do purchase intentions predict sales? *International Journal of Forecasting*, **23**, 347-364.

NIKOLOPOULOS, K., GOODWIN, P., PATELIS, A. & ASSIMAKOPOULOS, V. (2007) Forecasting with cue information: a comparison of multiple regression with alternative forecasting approaches. *European Journal of Operational Research*, **180**, 354-368.

SIMON, H. & SEBASTIAN K-H.(1987) Diffusion and advertising: the German telephone campaign. *Management Science*, **33**, 451-466.

SCHMITTLEIN, D.C. & MAHAJAN, V.(1982) Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing Science*, **1**, 57-78.

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

SRIVASTAVA, R.K., MAHAJAN, M., RAMASWAMI, S.R. & CHERIAN, J. (1985) A multiattribute diffusion model for forecasting the adoption of investment alternatives for consumers. *Technological Forecasting and Social Change*, **28**, 325-333.

TASHMAN, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, **16**, 437-450.

TAYLOR P.F & THOMAS, M.E. (1982) Short term forecasting: horses for courses. *Journal of the Operational Research Society*, **33**, 685-694.

THOMAS , R.J. (1985) Estimating market growth for new products; an analogical diffusion model approach. *Journal of Product Innovation Management*, **2**, 45-55.

TIGERT, D. & FARIVAR, B. (1981) The Bass new product growth model: a sensitivity analysis for a high technology product. *Journal of Marketing*, **45**, 81-90.

TULL, D.S. (1967) The relationship between actual and predicted sales in new product introduction. *Journal of Business*, **40**, 233-250.

VAN DEN BULTE, C., & LILIEN, G.L. (1997) Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Science*, **16**, 338-353.

WU, S.D., AYTAC, B., BERGER, R.T., & ARMBRUSTER, C.A. (2006) Managing short life-cycle technology products for Agere Systems. *Interfaces*, **36**, 234-247.