An Evolutionary Model of
New Product Tracking with Multiple Data Sets

Makoto Abe
Hotaka Katahira
▌Faculty of Economics, University of Tokyo ▌

ITME Discussion Paper No.79
May 2001
An Evolutionary Model of New Product Tracking
with Multiple Data Sets

Makoto Abe
abe@e.u-tokyo.ac.jp

Hotaka Katahira
katahira@e.u-tokyo.ac.jp

Faculty of Economics
University of Tokyo
Hongo, Tokyo
113-0033  JAPAN

April 2001
多種類データを用いた進化型新製品売上予測モデル

阿部 淑
片平 秀貴

東京大学大学院経済学研究科

製品のライフサイクルは近年ますます短くなる傾向にあり、新製品の売上をなるべく早い時点で予測することは非常に重要である。情報化社会において、企業、特にネット関連の会社は、売上、出荷、問い合わせなど新製品の売上評価に役立つような様々な取引データに瞬時にアクセスすることが可能である。既存の新製品売上予測モデルは多種のデータから情報を導き出すことに関しても、またマーケターのタイムリーなニーズに答えることに対しても、十分に対応出来ない状況にある。

既存の新製品売上予測モデルでは、提案された消費者のトライアル・リピート購買数理モデルのパラメータをその製品の初期販売データから必要な部分のみを使って推定し、他の部分は特に使われない。この研究では、（1）類似カテゴリーで過去に投入された新製品のデータと（2）市場調査会社が日常的に収集する多種類のデータを利用して、情報を最大限有効に使った新製品売上予測モデルを提案する。特性が未知な売上以外のデータからも売上予測に係関する情報を導き出すために、ノンパラメトリック手法が導入された。例えば、売上に関係したトライアル率の数理モデルひとつを拡げてみても、幾何関数、指数関数、エーランジメント k 関数、ワイプル関数、ガンマ-指数のような複合関数、ハザード関数など様々なパラメトリック関数が提案されており、個々のデータに適したパラメトリック関数を探索し正当化するプロセスを省くことは分析者の時間と労力の削減に大きく寄与する。このモデルは Little(1970) が提唱した意思決定サポート・システムの実用化に望ましい基準・単純、頑健、適応、進化を満たすため、実務での有用性が期待される。

モデルの検証は 1995 年から 1997 年に発売が開始された 13 のビール・ブランドのスキャナー・パネルデータを用いて行った。既存のモデルに基づいた基準モデルと比較して売上予測の精度が優れていた。本モデルは、データの収集期間が長いほど、過去の新製品の例が多いほど、そしてデータの種類が増えるほど予測精度が向上することからも進化型と言えるだろう。
ABSTRACT

While the effort to develop a new-product tracking model for consumer package goods --- one of the oldest methodologies in marketing science --- has reached its peak in mid 80s, the market and data environments have changed drastically these ten plus alpha years. Product life cycle has shortened and the needs for assessing the new product launch performance in the very early stage have become critical. Firms, especially e-commerce companies, now have an instant access to vast types of data regarding their past transactions, such as actual sales, shipment, inquiries, etc, which could potentially provide further information in evaluating the new product launch performance. Under such circumstances, the traditional tracking technologies seem very short of both exploiting the given informational frontiers and meeting the demands of the agile marketers today.

Despite the difference in sophistication, the existing models attempt to make forecast using only the data for the product of interest, and utilizing only the portion that is necessary to calibrate a postulated trial-repeat behavioral model and discarding the remaining information even if they are available as a part of routine data collection. The present model is proposed to bring the tracking technologies up-to-date by (1) incorporating data on previously introduced new products and (2) making use of all available information from multiple data sets that are collected routinely. A nonparametric projection method is introduced to extract tracking information from non-sales performance criteria whose properties are not well known. Even for trial rate projection alone, for example, past study suggested numerous parametric forms such as geometric, exponential, Erlang-k, Weibull, mixing distributions like exponential-gamma, and hazard models. A process of searching and justifying an appropriate parametric form for individual criteria can be bypassed, saving much of time and effort of the analyst. The proposed model is simple, robust, adaptive, and evolutionary --- desired implementation criteria for decision support models advocated by Little (1970).

The proposed model is tested on 13 new brands of beer introduced in Japan between 1995 and 1997 using scanner panel data and is shown to outperform the benchmark model based on the existing model. The model is evolutionary in the sense that the accuracy of forecasting improves as more periods, more examples of past new product launches, and more data sets become available.

Key Words: new product, consumer package goods, forecasting, nonparametric
1. INTRODUCTION

The new-product tracking model is one of the oldest methodologies in marketing science. Its origin traces back to the pioneering work by Parfitt and Collins (1968). The first comprehensive commercial model was developed by Blattberg and Golanty (1978), which came to be widely known as TRACKER. Although some variants have been proposed by the followers like NEWS (Pringle, Wilson and Brody 1982), the basic structure and philosophy of the model based on the buyers' trial and repeat behavior has remained more or less unchanged (Lilien, Kotler and Moorthy 1992).

Considering the fact that it is one of the most extensively and successfully applied areas of marketing science, it is hard to believe that the area has remained without significant innovations. Our conjecture is that most of the interesting ideas are kept proprietary due to the commercial nature of the applications and that they have never been integrated together and elaborated on to give rise to a significant extension of the model.

Despite slow progress in model building, the market and data environments have changed drastically these ten plus alpha years. The needs for assessing the new product launch performance in the very early stage have become critical. In the beverage category in Japan, for example, the retailers typically review the sales performance of newly introduced products and make an ONNO decision at week 12 or even at week 4. The decision is especially important for space limited convenience stores, whereby the largest chain, 7-Eleven, has recorded the highest sales in retailing industry, exceeding that of the largest supermarket chain, Daiei, for the first time (Nikkei Newspaper: April 12, 2001). It forces the marketers to become very agile and dynamic in running the follow-up marketing programs. Consequently, the demand for more accurate assessment at the very early stage of a launch has become critical.

Such agile market environments together with the progress of supporting information technologies have given rise to surprisingly rich data environments. At one extreme, an e-commerce company would enjoy perfectly accurate sales data at every minute of their operation. They would have an instant access to almost any types of data regarding their past transactions, so far as their database is properly maintained. Such data might include, in addition to the actual sales, shipment, inquiries, non EC sales, etc., which could potentially provide further information for assessing the new product launch performance. Under such
circumstances, the traditional tracking technologies seem very short of both exploiting the given informational frontiers and meeting the demands of the agile marketers today. The present model is proposed to bring the tracking technologies up-to-date by making the model evolutionary and compatible with multiple data set environments.

Firstly, the model is evolutionary. The proposed model not only uses the current data for the launched brand in question but also incorporates that of all past launch histories in a relevant category. This would make the model more accurate the more often it was used in the new product launch context in a given category.

Secondly, it is capable of integrating multiple data sets. The existing tracking models typically postulate a behavior model incorporating, for example, awareness, trial and repeat purchases, and calibrate on panel data collected by survey in conjunction with (pre-) test market data. In the real launch contexts even in the traditional (non-EC) retailing, however, other relevant data like retail shipping data and store POS data, which may not be necessary to calibrate such behavioral models but still provide useful insights, usually exist. If (pre-) test marketing and/or longitudinal marketing research (survey) are conducted, then additional data such as product awareness, stated-preference, and satisfaction in time-series might be available. What is being done, if at all, under such circumstances is to make separate predictions and diagnosis using each data set one by one and to make 'manual' reconciliation to arrive at the final conclusion. Obviously, it would contribute to the accuracy of predictions if there were any way of systematically integrating the multiple data sets available (Russell and Kamakura 1994). The present paper only demonstrates how such problem is well taken care of in our model framework and leaves the empirical investigation to a separate paper.

In the next section, the literature in the field was briefly reviewed and their problems in applications are identified. The model is specified in section 3 and empirical results are presented in section 4. Summary and conclusion are given in the final section.

2. REVIEW OF PREVIOUS MODELS

New product forecasting has been a popular research area in Marketing Science in the 70s and 80s. One major research stream applies to frequently purchased consumer goods that involve trial and repeat purchases. The initial new product forecasting model of trial-repeat traces back to the pioneering work by Parfitt and Collins (1968), whereby a market share is
determined by three factors: penetration rate, repeat rate and buying rate. More sophisticated models were introduced since then, and some were implemented by marketing research firms who applied them to real data. For this reason, many of the well known models possess a trademark such as SPRINTER (Urban 1970), TRACKER (Blattberg and Golanty 1978), ASSESSOR (Silk and Urban 1978), and NEWS (Pringle, Wilson and Brody 1982). We suspect that many more are kept proprietary from academic literature.

Ideas behind these methods are basically the same in that a trial and repeat behavioral model is postulated, some incorporating marketing mix effect such as pricing, advertising, promotion, and distribution, and its parameters are calibrated on trial and repeat data from (pre-) test market and survey data. For example, in TRACKER, ASSESSOR, and NEWS, repeat propensity and usage rate are solicited by survey to complement sales data observed in (pre-) test marketing. Product awareness is gathered by monthly survey in TRACKER and NEWS, and in conjunction with actual GRP aired, effect of advertising response is evaluated. Effectiveness of sample drop is estimated from follow-up telephone call in ASSESSOR.

Since early 80s, however, no significant progress has been made to trial-repeat models, at least in the academic literature. NEWS was the very first article in the first issue of Marketing Science (1982). Yet, a special issue on new product development in the Journal of Marketing Research (February 1997) does not contain even a single article on a forecasting model of frequently purchased goods. An academic focus has shifted to another stream of research in new product forecasting, characterized by diffusion models. Originally introduced by Bass (1969), these models apply mainly to durable goods and technology where the focus is to predict adoption of innovation.

Some tried to extend diffusion models to a repeat purchase setting by combining a Markov model that describes movement of buyers in different purchase states such as non-trier, trier, repeater and past user (Dodson and Muller 1978, Hahn, Park, Krishnamurthi, and Zoltners 1994, Lilien, Rao, and Kalish 1981). Diffusion models are basically macro flow models whose parameters are calibrated from aggregated data. These, so called repeat purchase diffusion models, are no exceptions. Based on aggregate data, a postulated model dictates how sales are decomposed into two components, one due to trial and another due to repeat. This is in contrast to trial-repeat type models that start from a buyer behavior process whose parameters such as trial, repeat and usage rates are calibrated separately using appropriate information from panel data.
The proposed forecasting model follows the spirit of the former trial-repeat type, but presents two major distinctions from the existing models such as SPRINTER, TRACKER, ASSESSOR, and NEWS.

First, despite the difference in sophistication, the existing trial-repeat forecasting models attempt to make forecast using only the data for the product of interest. Recent advance in information technology makes it relatively easy to collect panel data through UPC scanner or over internet, and to create knowledge base by accumulating past cases of new product launches in a history database. In such information era, a forecasting model that can incorporate data on previously introduced new products should possess a distinct advantage.

Second, the existing models postulate a behavioral model of a trial-repeat purchases, expend a lot of effort justifying the model through statistical testing, then calibrating its parameters with data (Fader, Hardie and Zeithammer 1999). If routinely collected data from (pre-) test market do not provide sufficient information for estimating such a model, additional data are collected by customized survey. On the other hand, data that are not needed to calibrate the postulated model are simply discarded even if they are available as a part of routine data collection.

We do not think building a more sophisticated behavior model is the right approach. One can always introduce complications like a time-varying trial rate, or repeat and usage rates that vary by repeat depth. Such models often require specialized data, and a large number of parameters cannot be estimated reliably, thereby lacking robustness. Even worse, values of certain parameters must rely on subjective judgment. That was the case in TRACKER for the repeat rates of different depths, d(k).

In this information intensive era, we propose a method that shifts the emphasis from model to data. We seek any information that might help in assessing the performance of a new product launch from routinely collected data by research firms alone. Data might be available from multiple sources such as panel data, store POS data, retail shipping data, (pre-) test marketing, and marketing research/survey. We then construct a series of simple and sound models, behavioral and non-behavioral, to these pieces of information. Behavioral models should link the input information to sales assuming only basic behavioral theories, whereas non-behavioral models can be constructed based either on sound theories or a simple mathematical relationship that links the input information to sales. As shown in the
application, even a nonparametric mathematical relationship can be employed to avoid adverse influence of an improper parametric assumption or to bypass the lengthy process of selecting an appropriate parametric form altogether.

To summarize, the proposed new product forecasting model makes use of all available information that is collected routinely by research firms. In particular, the model incorporates data on (1) previously introduced new products and (2) non-sales performance criteria from different data sources, to improve forecasting accuracy. It can be applied easily to a real setting as a forecasting and/or an early warning / diagnostics system.

3. MODEL

The objective of the proposed model is to forecast sales of a new consumer package product at the early stage of its introduction. In particular, many marketing managers would like to know one-year ahead (i.e., 52nd week) cumulative sales prediction at the nth week from the launch, where n is as early as 12, for planning manufacturing and reviewing the marketing program.

The basic idea of the model is as follows. Multiple performance criteria (sales and non-sales) obtained from various data sets, such as scanner panel data, store POS data, retail shipment data, are used to compare the new product with previously introduced products in similar categories during the first n weeks to identify products that exhibit similar growth patterns thus far. Based on the 52nd week performance of these products, we make a 52nd week projection of the given product with respect to each criterion, which is the n converted to a sales projection. Finally, multiple sales projections obtained from different performance criteria are reconciled. The step-by-step procedure is illustrated in Figure 1 and described below in detail.

< Figure 1: Procedure of the Model >

In what follows, we limit our attention to typical scanner panel data routinely provided by a marketing research firm. However, other data sources, such as retail shipment records and time-series survey results, can be incorporated to further improve the forecasting accuracy. The data used in this research come from scanner panel data called Quick Purchase Report of Tokyu Agency (http://www.tokyu-agc.co.jp/QPR/) --- a standard market report provided to its client manufacturers. They contain aggregated weekly statistics of consumer package
products such as sales, number of purchases, penetration rate, and repeat rates for depth of up to 4th times. Non-sales criteria used in this research are cumulative number of purchases, a penetration rate, and a repeat rate. Also assume that sales and number of purchases are normalized to a unit of per 100 households, as is the case for QPR data.

**Step 1: Estimating the 52nd week performance criterion from past new products that exhibit "similar growth pattern" during the first n weeks.**

We compare the new product in question with previously introduced products in the same or similar categories on various performance criteria, including not only cumulative sales but also others such as the number of purchases and penetration and repeat rates during the first n weeks. For each criterion, we make the 52nd week prediction based on the 52nd week performance of those products that exhibit "similar" growth patterns during the first n weeks.

To quantify "similarity in growth pattern", the time-series plot of a chosen performance criterion (e.g., cumulative sales, penetration, etc.) is normalized so that all products intersect at the nth week. This is to eliminate the scaling difference across products and to extract only growth patterns. Similarity is measured by the likelihood (probability) that the given product belongs to any of the past products in the history database, assuming a normal distribution around the observed value. We then weight the 52nd week performance figures of the past new products by this likelihood to come up with the 52nd week prediction for the product.

To clarify the process, consider an example in which there are 10 past new products in the database. Then, depending on how similar their growth patterns are to the product under consideration during the first n weeks, a probability is assigned to each of the 10 products. These probabilities sum to one. Then, the 52nd week performance figures of the 10 products are added with weights equal to these probabilities to come up with the 52nd week prediction for the product.

**Step 2: Estimating the relationship between sales and non-sales performance criteria**

We establish the relationship between sales and these non-sales criteria for the product using the first n weeks of data. In other words, for each performance criterion, we estimate a function $f$, where $sales = f(criterion)$.

$$Cumulative Sales = f(Cumulative Number of Purchases)$$
While quantity bought at each purchase occasion might vary depending on consumers and whether it is a trial or repeat purchase, on average, the following simple relationship exists.

\[ Sales(t) = (\text{Quantity per Purchase}) \times (\text{Number of Purchases}) \]

By regressing cumulative sales on the cumulative number of purchases without an intercept using the first n weeks of data, we can estimate the "Quantity per Purchase", and thus obtain the above relationship.

**Cumulative Sales = \( f (\text{Penetration, Repeats}) \)**

Consider more elaborate model than the above relationship, whereby different usage rates apply to triers and repeaters. Then it is shown in the appendix that the following relation holds.

\[ \text{CumulativeSales}(t) = \alpha_0 \times R_0(t) + \alpha_1 \times R_1(t) \]

where \( R_0(t) \) and \( R_1(t) \) are penetration and repeat rates, respectively, and \( \alpha_0 \) and \( \alpha_1 \) are multiplicative factors derived in the appendix. Again, by regressing cumulative sales on penetration and repeat rates using the first n weeks of data, we can estimate \( \alpha_0 \) and \( \alpha_1 \).

If other performance criteria are used, their relationship to sales must be obtained in a similar manner. In our case, \( f(\cdot) \) had a clear relationship from a theoretical standpoint. If an appropriate theory or logic does not exist, however, a purely mathematical relationship (using some sort of regression) can be established.

**Step 3: Converting the 52nd week projection of non-sales criteria to sales forecast**

The 52nd week projections for the number of purchases, penetration rate, and repeat rate obtained in step 1 are converted to the corresponding sales figure using the relationship established in step 2. The process results in three 52nd week sales predictions based on (1) the first n weeks of sales, (2) the number of purchases, and (3) penetration and repeat rates.

**Step 4: Combining sales predictions obtained from different performance criteria**

We must combine the three sales predictions (1) ~ (3) to come up with a single prediction. Each sales prediction underlies a different behavioral assumption. Prediction (3) involves
the most elaborate behavior model by assuming different usage rates between triers and repeaters, whereas prediction (2) do not make such distinction between these two groups of buyers. Prediction (1) skips a behavior model all together and regard sales simply as a number. Therefore, internal validity increases from criterion (1) to (3) as more elaborate behavior models are involved, whereas external validity is decreasing since the performance criterion becomes a lesser direct measure of sales. With such a tradeoff, it appears difficult to judge a priori which criterion produces the most reliable sales prediction. An approach we propose, in the absence of a priori expectation, is to give equal weights on all criteria by taking an average of the three predictions. Averaging should smooth out any idiosyncratic noise associated with particular performance criteria.

Two distinguishing features of the proposed forecasting model are,

**Evolutionary:** Because past new products are benchmarked, there exists a higher chance of identifying a similar growth pattern as more products are added to this history database, which in turn results in an improved forecast.

**Multiple Data Sets:** As multiple performance criteria are employed, it is expected that forecasting accuracy improves with the number of criteria being used, and hence, the number of data sources.

Because scanner panel data did not contain information on marketing activities of the product under consideration and those of competitors, the current model did not address the effect of marketing mix explicitly. While it is possible to extend our model to include the marketing mix effect, real difficulty lies in obtaining competitive rather than its own marketing activities when those past new products in the history database were launched.

4. **EMPIRICAL RESULTS**

In this section, an empirical application of the model is illustrated.

4.1. **Data**

The QPR data provided by Tokyu Agency contained panel data on 13 new brands of beer introduced between 1995 and 1997. The information was collected by home scanning from a panel of 2500 households who live in the Tokyo metropolitan area. The data provided aggregated weekly statistics such as sales, number of purchases, penetration rate, and repeat
rates of up to 4th times, for 52 weeks since their launch.

Our empirical analysis proceeds as follows. We pick a brand to forecast and assumed that the sales data of only the first n weeks are known. The remaining 12 brands are regarded as previously introduce brands whose sales data are known for the entire 52 weeks in the history database. For different calibration lengths (i.e., n = 4, 8, 12, 16, 24, 48 weeks), we forecast the brand's sales up to the 52nd week. A forecasting brand is then rotated.

4.2. Step 1: Estimating the 52nd week Performance Criterion

The basic idea here is to combine cross-sectional data from other products to complement its own time-series data. Along this idea, two methods, parametric and nonparametric, are proposed. The nonparametric method (we will refer to it as NM-Track, which stands for nonparametric multiple tracking) assigns a likelihood (probability) to each of the 12 past new products, depending on how similar the growth pattern is during the calibration period. Forecast beyond the calibration period is obtained by weighting the performance of the past products with this likelihood. The procedure can be interpreted as an interpolation of cross-sectional data, and thus it has an advantage that the forecast cannot go wild, overshooting or undershooting beyond the past range.

Paralleling the procedure, its parametric counterpart (we will refer to it as PM-Track, which stands for parametric multiple tracking) is as follows. First, an appropriate parametric function is fit to the time-series plot for each of the past products. Then a likelihood of the brand's performance to each of the 12 fitted parametric curves is computed, depending on how similar the growth pattern is during the calibration period. Then a forecast is obtained by weighting these 12 parametric curves according to the likelihood. Because the parametric method involves curve fitting, it is not as flexible as the nonparametric counterpart. With appropriate modification, however, the parametric method can overcome a difficulty of the nonparametric method -- namely inability to extrapolate. The modification is as follows.

The 12 parametric curves are reduced to a few "typical" curves, say five that represent High (H), Medium High (MH), Medium (M), Medium Low (ML), and Low (L) growth rates. To permit extrapolation beyond the past range, H is chosen to be higher than the historical maximum, say by 20%, and L is chosen to be lower than the historical minimum, again by 20%. By doing so, the method can accommodate forecasting of a product that exhibits better or worse performance thus far than any of the past products in the history database (up to
20%). Because the nonparametric method offers flexibility but extrapolation, and vice versa for the parametric method, NM-Track and PM-Track should complement the forecast.

We compare these nonparametric and parametric methods with a naive parametric curve fitting method that does not incorporate cross-sectional data. With the naive method, a parametric function is fit to the calibration data of a given brand, and information on previously launched products in the history database is not used. Existing models such as SPRINT, TRACKER, ASSESSOR, and NEWS, while incorporating marketing mix or modeling an awareness stage and varying usage for different depths in repeat purchase, all attempt to make forecast only from the data for the product of interest. Because syndicated scanner data provided by marketing research firms usually do not contain information on advertising GRP, brand awareness or product usage, these models, if implemented on our routinely collected scanner data, reduce to this naive method. Thus, the naive method would be a benchmark for the existing forecasting models. Subsequently, we report the comparison of the three methods on each of the four performance criteria: cumulative sales, cumulative number of purchases, penetration rate, and repeat rate.

Figures 2, 3, and 4 presents forecasting of cumulative sales for the naive, parametric, and nonparametric methods, respectively, with 12 weeks of a calibration period (shown as a vertical line). For the naive and parametric methods of Figures 2 and 3, the best fit was obtained with \( f(t) = \alpha(1 - e^{\beta t}) \) after trying out several functions. Figure 2 shows overshooting and undershooting for some brands, resulting in a large error toward the 52nd week. The difficulty of making a one-year ahead forecast at the 12th week is quite obvious. For each brand, the left plot of Figure 3 shows the 5 guidelines (H, MH, M, ML, L) obtained from the past new products and the right plot shows the forecast along with the observed sales. Figure 4 shows the sales of the 12 past brands on the left and the forecast on the right along with the actual sales.

< Figure 2: Forecast of Sales by the Naïve Method >

< Figure 3: Forecast of Sales by the PM-Track >

< Figure 4: Forecast of Sales by the NM-Track >

Accuracy of the forecast is reported in Figure 5, in which the prediction error associated with each method in terms of mean absolute fractional error (MAE) during the forecasting period
is plotted for various lengths of a calibration period. The error is smaller with the nonparametric method than with the parametric method, and the naive method has the largest error as expected for all calibration lengths except 48 weeks.

< Figure 5: Mean Absolute Fractional Error of the Forecast of Sales >

We now turn to the forecast of penetration. Figure 6 shows the penetration rate of each brand, showing a different scale and a growth pattern. As for cumulative sales, $f(t)=\alpha(1 - e^{-\beta t})$ resulted in the best fit among several functions. Figure 7 reports the MAE for each method with different lengths of a calibration period. Here, the advantage of the nonparametric method is most prominent at the early stage of forecasting like the 4th or 8th week, since no parameters need to be estimated from a small amount of data.

< Figure 6: Time-Series Plot of Penetration >

< Figure 7: Mean Absolute Fractional Error of the Forecast of Penetration Rate >

Figures 8 and 9 show the corresponding MAEs for a repeat rate and a cumulative number of purchases, respectively. For the naive and parametric methods, again the same function, $f(t)=\alpha(1 - e^{-\beta t})$, had the best fit and was used in both performance criteria.

< Figure 8: Mean Absolute Fractional Error of the Forecast of Repeat Rate >

< Figure 9: Mean Absolute Fractional Error of the Forecast of Number of Purchases >

Summarizing these results, the following observations can be made. (1) Regardless of the performance criterion or calibration length, the error is the smallest with the nonparametric method, followed by the parametric method, and then by the naive method. (2) Advantage of the nonparametric method is more prominent for shorter calibration lengths. (3) As the calibration length increases, the difference in forecasting performance among the three methods decreases.

4.3. Step 2: Establishing the Relationship between Sales and Non-Sales Criteria

The relationship between sales and the number of purchases is

$$ Sales(t) = \text{Quantity per Purchase} \times \text{(Number of Purchases)}.$$
Cumulative sales is regressed on the cumulative number of purchases using data for a calibration period. Because quantity per purchase differ greatly by products, a separate regression is run for each brand. All regression had good fit with the average $R^2$ of over 0.98. The average quantity per purchase across brands was 1724 ml or about 5 cans.

The relationship between sales and penetration / repeat rates is

$$\text{Cumulative Sales}(t) = \alpha_0 \times R0(t) + \alpha_1 \times R1(t).$$

Cumulative sales is regressed on penetration and repeat rates using data from a calibration period. Again, all regression had good fit with the average $R^2$ of over 0.98.

Hence, the relationship between sales and non-sales criteria is quite strong.

### 4.4. Step 3: Obtaining Sales Prediction from Each Performance Criterion

Figure 10 compares four sales predictions based on (1) sales from the first n weeks benchmarking the past new products in step 1, (2) the number of purchases, (3) penetration and repeat rates, (4) a combination of the first three predictions through averaging. Figure 11 and Table 1 report the MAE for these four predictions (1) ~ (4) along with a prediction by the naïve method, which does not incorporate the data on past products.

< Figure 10: Sales Prediction based on Different Performance Criteria >

< Figure 11: Mean Absolute Fractional Error of the Forecast of Sales >

< Table 1: Mean Absolute Fractional Error of the Sales Forecast by Different Criteria >

When the information from the past products is incorporated, one can clearly see the improvement in forecast as evidenced by the smaller MAE of methods (1) ~ (4) over the naïve method. Depending on the length of a calibration period, a performance criterion that produces the most accurate forecast differs. A general tendency is, however, use of behavioral criteria such as the number of purchases of (2) and penetration and repeat of (3) seems to help more in early stages, whereas direct sales of (1) is more useful toward the end. For example, use of penetration and repeat results in the lowest MAE for the 4 and 8 week calibration periods, whereas use of sales shows the least error for the 24 and 48 week calibration periods. This observation could be attributed to the tradeoff between internal and external validities of these criteria.
4.5. Step 4: Reconciling Sales Predictions obtained from Various Performance Criteria

Under the absence of a clear dominant criterion and no prior preference to a particular criterion, a first-cut approach might be to take their average. That should cancel out noise associated with any particular method for any particular calibration period. It appears that the combination method of (4), insensitive to the calibration length, produces a robust and stable forecast than singling out any particular criterion.

5. SUMMARY

With so many new products introduced every months, it is crucial to forecast a new product’s performance at the earliest stage and apply necessary marketing tactics. In particular, many marketing managers would like the 52nd-week sales prediction figure as early as 12 weeks into its introduction. Most current approaches involve postulating a trial-and-repeat behavior model and calibrating its parameters using the first few months of data. With only a handful of data points, however, unexpected parameter estimate can often result. To improve forecasting accuracy, we incorporate data on (1) previously introduced new products and (2) non-sales performance criteria from difference data sets — two distinct features of the proposed model.

Despite the difference in sophistication, the existing study on new product forecasting for trial-repeat goods attempt to make prediction using only the data for the product of interest. Furthermore, they do not make full use of routinely collected data: utilizing only a portion of data that is necessary to calibrate the postulated behavioral model and discarding the remaining information that is not needed for the calibration. In this sense, if implemented on our scanner data collected routinely by a research firm, these models reduce to our naïve method we benchmarked against. We tested the model on 13 new brands of beer introduced between 1995 and 1997 using scanner panel data.

(Step 1) Two methods, the nonparametric and parametric methods, were compared against the naïve method -- a simple time-series projection without using data on past new products. The empirical results showed that, in all four performance criteria, the forecast was most accurate with the nonparametric method, then the parametric method, followed by the naïve method. The advantage of the nonparametric method was most prominent for shorter calibration periods. As the calibration period lengthened, the mean prediction error diminished in all methods, and the difference in forecasting accuracy among the methods
decreased.

(Step 2) The relationships between sales and non-sales performance criteria were derived on the basis of theoretical consideration shown in the appendix, and their parameters were estimated by regression. All relationships appeared to be quite sound, resulting in the average $R^2$ of 0.98.

(Step 3) Regardless of the performance criterion used, sales predictions by the nonparametric method incorporating information from past products outperformed the forecast by the naïve method. No single performance criterion dominated the forecast. Use of behavioral criteria like penetration and repeat rates or the number of purchases seemed to help more in early stages, whereas direct sales was more useful toward the end.

(Step 4) The average of sales predictions that were obtained using different criteria produced the most robust and stable forecast, insensitive to the length of a calibration period.

The proposed model is simple, robust, adaptive, and evolutionary --- desired implementation criteria for decision support models advocated by Little (1970). The parametric forms used in the model are simple and hence produce a robust estimate. The idea behind the nonparametric method of the model is straightforward and results in a robust forecast thorough interpolation. The model is evolutionary in the sense that the prediction improves as more periods, more historical examples of new products, more performance criteria, and more data sets become available.

Use of nonparametric functions in Step 1 is especially convenient when there are many performance criteria that need to be incorporated in forecasting. A process of searching and justifying an appropriate parametric form for individual criteria can be bypassed, saving much of time and effort of the analyst. Even for trial purchase projection alone, for example, past study suggested numerous parametric forms such as geometric (Fourt and Woodlock 1960, TRACKER, ASSESSOR), exponential (Parfit and Collins 1968), Erlang-k, Weibull (Massy, Montgomery and Morrison 1970), mixing distributions like exponential-gamma, and hazard models (Helsen and Schmitttlein 1993), to name a few (see Fader, Hardie and Zeithammer 1999 for extensive comparison). The nonparametric method can make the analysis more or less auto-pilot, which is especially suitable to the data-abundant IT environment.

A Bayesian approach to new product forecasting also uses data on past products to
compensate for a short period of data (Lilien, Rao, Kalish 1981, Sultan, Farley, Lehmann 1990). However, there exists a crucial difference between our use and the Bayesian use of historical data. In the Bayesian approach, past data is used to estimate a prior distribution of model parameters. That is, to set a prior value to "norm". Unfortunately, such a "norm" prior tends to pull (or shrink) the forecast toward an "average" prediction. Experience with real data suggests that each product is so different that an "average" input generating an "average" result is not particularly insightful. In contrast, recognizing the difference of individual products, our model identifies a few products from the past that match their idiosyncrasies. Thus, the concept is fundamentally different from the Bayesian approach. In fact, our model can add the Bayesian flavor by introducing a prior as to which past products are most similar to the current product when evaluating the likelihood. The prior can come from manager’s experience and intuition, and results based on concept and proto-type testing, test marketing, and one-time survey.

Let us make a note on limitations as well as directions for extending the current research. First, the empirical result is based on one category, and they may not generalize to other categories. In particular, the superiority of the nonparametric over the parametric method may not apply if a product exhibits an unusual growth pattern that was not observed in the past. The parametric method, although not as flexible as the nonparametric method in fitting various shapes of time-series trajectories, can predict beyond the range of the past products through extrapolation as was done in our application. In such a case, the parametric method might perform better. Hence, the parametric and nonparametric methods should complement each other in obtaining more accurate forecast. It certainly adds to our understanding if the model is tested with more product categories.

Second, we did not observe a single dominant performance criterion that leads to a superior prediction, and the simple averaging resulted in the most robust prediction. However, sales predictions based on different criteria could be combined with weights. One useful extension is to obtain such weights rather than putting equal weights.

Third, the proposed model does not incorporate the effect of marketing mix. The decision was made to eliminate additional data collection beyond those available from standardized scanner panel data provided by a marketing research firm. For example, to model the effect of advertising on awareness, one must know aired advertising intensity (in GRP perhaps) and conduct survey on product awareness. Even if such data are available, they tend to be
contaminated by marketing activities of competitive products, and managers rarely have
detailed knowledge on such competitive information. In fact, none of the forecasting models
we reviewed here accounts for such competitive effect. For simplicity and robustness, we
decided not to include the effect of marketing mix. To avoid modeling a complication
arising from marketing activities of incumbent brands, the sales of past new launches should
account for such a second order competitive effect and the corrected sales figure should be
compiled in the history database. If the data is available, however, that would be another
fruitful extension to the model.

APPENDIX

Analytical Relationship between Cumulative Sales and Penetration and Repeat Rates

Define a penetration rate at week t, R0(t), as a fraction of households who purchased at least
once. Likewise define a repeat rate of depth one, R1(t), as a fraction of households who
repeated at least once (i.e., who purchased at least twice). In general, a repeat rate of depth n
is a fraction of households who repeated at least j times and can be denoted as Rj(t).

First, observe that the following identify.

\[
\sum_{j=0}^{\infty} R_j(t) = \sum_{j=0}^{\infty} \text{Cumulative Number of Purchases}(t) = \sum_{j=0}^{\infty} \text{Cumulative Sales}(t) = \sum_{j=0}^{\infty} R_j(t) \cdot UR_j
\]

Let us denote the usage rate for the j-th repeater as URj. Then,

\[
\text{Cumulative Sales}(t) = \sum_{j=0}^{\infty} UR_j \cdot R_j(t)
\]

We now make two simplifying assumptions. First, usage rates for repeaters are the same
regardless of the depth. That is UR1 = UR2 = … = URj = … Usage rates for triers and
repeaters can now be denoted as UR0 and UR1, respectively. Second, a repeat rate of a
deeper level is some fraction of the first repeat rate. That is Rj(t) = βj × R1(t) for j=2,3,…
This is justified because the linear relationship holds with high R^2 (over 0.95) when
regression is run on the raw data. Applying these simplifications results in the desired
relationship between cumulative sales and penetration / repeat rates.
Cumulative Sales(t) = UR0 R0(t) + UR1 R1(t) \left( 1 + \sum_{j=2}^{\infty} \beta_j \right) \\
= \alpha_0 \times R0(t) + \alpha_1 \times R1(t)

REFERENCES


Table 1: Mean Absolute Fractional Error of the Sales Forecast by Different Criteria

<table>
<thead>
<tr>
<th>week</th>
<th>Sales Naive</th>
<th>Sales NM-Track</th>
<th>Number of Purchases</th>
<th>Penetration &amp; Repeat</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.6376</td>
<td>0.4991</td>
<td>0.3125</td>
<td>0.2538</td>
<td>0.3226</td>
</tr>
<tr>
<td>8</td>
<td>0.4621</td>
<td>0.2920</td>
<td>0.2574</td>
<td>0.2068</td>
<td>0.2334</td>
</tr>
<tr>
<td>12</td>
<td>0.2827</td>
<td>0.2205</td>
<td>0.1947</td>
<td>0.1952</td>
<td>0.1762</td>
</tr>
<tr>
<td>16</td>
<td>0.2683</td>
<td>0.1962</td>
<td>0.1632</td>
<td>0.1731</td>
<td>0.1633</td>
</tr>
<tr>
<td>24</td>
<td>0.2278</td>
<td>0.1251</td>
<td>0.1511</td>
<td>0.1602</td>
<td>0.1238</td>
</tr>
<tr>
<td>48</td>
<td>0.0310</td>
<td>0.0175</td>
<td>0.0274</td>
<td>0.0449</td>
<td>0.0249</td>
</tr>
</tbody>
</table>

Figure 1: Procedure of the Model
Figure 2: Forecast of Sales by the Naïve Method

- SP Drafty
- K Half&Half
- K Jiyuujiikan
- K LA2.5
- K Shokunin
- SU SuperHops
- SU Half&Half
- SU Daichi
Figure 3: Forecast of Sales by the PM-Track

SP Drafty

K Half&Half

K Jiyuujikan

K LA2.5
Figure 4: Forecast of Sales by the NM-Track

**SP Drafty**

- **Normalized Sales**
  - Week: 0 to 60
  - Y-axis: 0 to 5

**K Half&Half**

- **Normalized Sales**
  - Week: 0 to 60
  - Y-axis: 0 to 5

**K Jiyuuikan**

- **Normalized Sales**
  - Week: 0 to 60
  - Y-axis: 0 to 5

**K LA2.5**

- **Normalized Sales**
  - Week: 0 to 60
  - Y-axis: 0 to 5

**SP Drafty**

- **Sales**
  - Week: 0 to 60
  - Y-axis: 0 to $10^5$

**K Half&Half**

- **Sales**
  - Week: 0 to 60
  - Y-axis: 0 to $10^4$

**K Jiyuuikan**

- **Sales**
  - Week: 0 to 60
  - Y-axis: 0 to $10^4$

**K LA2.5**

- **Sales**
  - Week: 0 to 60
  - Y-axis: 0 to $10^4$
Figure 5: Mean Absolute Fractional Error of the Forecast of Sales

Figure 6: Time-Series Plot of Penetration
Figure 7: Mean Absolute Fractional Error of the Forecast of Penetration Rate

Figure 8: Mean Absolute Fractional Error of the Forecast of Repeat Rate
Figure 9: Mean Absolute Fractional Error of the Forecast of Number of Purchases
Figure 10: Sales Forecast based on Different Performance Criteria

SP Drafty

Using Sales

Using No. of Purchases

Using Combination

Using Penetration & Repeat

K Half&Half

Using Sales

Using No. of Purchases

Using Combination

Using Penetration & Repeat
The diagram shows four sets of graphs comparing sales for different models and strategies:

1. **K Shokunin**
   - Using Sales
   - Using No. of Purchases
   - Using Combination
   - Using Penetration & Repeat

2. **SU SuperHops**
   - Using Sales
   - Using No. of Purchases
   - Using Combination
   - Using Penetration & Repeat

The y-axis represents sales, and the x-axis represents a range from 0 to 40. The graphs display trends for each model and strategy.
Figure 11: Mean Absolute Fractional Error of the Sales Forecast by Different Criteria

MAE for Sales Forecast

- • Sales Naive
- ■ NM-Track
- ▲ Number of Purchases
- • Penetration & Repeat
- • Combination