# Forecasting in Marketing*

Philip Hans Franses†

*Econometric Institute*

*Department of Business Economics*

*Erasmus University Rotterdam*

Econometric Institute Report EI 2004-40

August 2004

**Abstract**

With the advent of advanced data collection techniques, there is an increased interest in using econometric models to support decisions in marketing. Due to the sometimes specific nature of variables in marketing, the discipline uses econometric models that are rarely, if ever, used elsewhere. This chapter deals with techniques to derive forecasts from these models. Due to the intrinsic non-linear nature of these models, these techniques draw heavily on simulation techniques.

Key words and phrases: Forecasting, Marketing, Koyck model, Bass model, Attraction model, Unobserved heterogeneity

*This first version of this chapter was prepared for inclusion in the Handbook of Economic Forecasting, to be edited by Graham Elliott, Clive Granger and Alan Timmermann, and to be published by Elsevier North Holland. I thank participants at the Rady Handbook of Economic Forecasting Conference (San Diego, April 2004), and Gerard Tellis, Richard Paap, Dennis Fok, and Csilla Horvath for many helpful suggestions. Lotje Kruithof provided excellent research assistance. The address for correspondence is Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, email: franses@few.eur.nl

†Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR, Rotterdam, The Netherlands, `email: franses@few.eur.nl`

# Contents

# 1 Introduction

In their recent bestseller, Kotler *et al.* (2002, page x) state that "Today's businesses must strive to satisfy customers' needs in the most convenient way, minimizing the time and energy that consumers spend in searching for, ordering, and receiving goods and services". Obviously, these authors see an important role for marketing activities to support that objective.

At the same time this statement indicates that marketing activities can be targeted at the level of an individual consumer's level, and that time is an important factor. Time can consider the speed at which consumers can respond, but it also concerns the ability to evaluate the success or failure of marketing activities. For a quick evaluation, one benefits from detailed data, observed at a high frequency, and preferably including performance data of competitors. With the advent of advanced data collection techniques, optic scanner data and web-based surveys, today's decisions on the relevant marketing activities can be supported by econometric models that carefully summarize the data. These basically concern links between performance measures as sales with marketing input like prices and advertising. Direct mailings for example can now be targeted at specific individuals, bonus offers in retail stores can be given to only a selected set of consumers, and the shelf position of certain brands is chosen with meticulous precision.

One of the academic challenges in this area is to design econometric models that adequately summarize the marketing data and that also yield useful forecasts, which in turn can be used to support decision-making[1]. The last few decades have witnessed the development of models that serve particular purposes in this area, and this chapter will describe several of these. The second feature of this chapter is to demonstrate how forecasts from these models can be derived. Interestingly, many of these models are intrinsically non-linear, and as will be seen below, so simulation-based techniques become mandatory.

The outline of this chapter is as follows. Section 2 briefly reviews the type of

[1]This chapter will be dedicated to models and how to derive forecasts from these models. The implementation of these forecasts into decision-making strategies is beyond the scope of this chapter, see Franses (2005a,b) for further discussion.

measures that are typically used to evaluate the performance of marketing efforts. These performance measures are sales, market shares, purchases, choice and time between events[2]. These variables are the outcomes of marketing activities that can concern pricing strategies, promotional activities, advertising, new product introduction, but can also concern the consequences of competitors' actions. Section 3 discusses a few models that are typically used in marketing, and less so, if at all, in other disciplines. Section 4 demonstrates how forecasts from these models can be generated. This section adds to the marketing literature, where one often neglects the non-linear structure of the models. Section 5 concludes this chapter with a few further research topics. The aim of this chapter is to demonstrate that there is an interesting range of econometric models used in marketing, which deserves future attention by applied econometricians and forecasters.

## 2 Performance measures

One of the challenging aspects of marketing performance data is that they rarely can be treated as continuous and distributed as conditionally (log) normal. Perhaps sales, when measured as quantity purchased times actual price, can be assumed to fit the classical assumptions of the regression model, but sales measured in units might sometimes be better analyzed using a count data model. Other examples of performance measures are market shares, with the property that they sum to 1 and are always in between 0 and 1, and the amount or the percentage of individuals who have adopted a new product. This adoption variable is also bounded from below and from above (assuming a single adoption per consumer). One can also obtain data on whether an individual makes a purchase or not, hence a binomial variable, or on whether s/he makes a choice amongst a range of possible products or brands (multinomial data). Surveys using questionnaires can result in data that are multinomial but ordered, like ranging from "strongly disagree" to "strongly agree"

[2] This chapter abstains for a discussion of how conjoint analysis, where stated preferences for hypothetical products are measured, can help to forecast revealed preferences measuring actual sales or adoption. This is due to the fact that the author simply has not enough experience with the material

on for example a 5-point scale. Finally, there are marketing data available which measure the time between two events, like referrals to advertising cues or, again, purchases.[3]

## 2.1 How do typical data sets look like?

To narrow focus towards the models to be reviewed in the next section, consider a few typical data sets that one can analyze in marketing.

**Sales**

Sales data can appear as weekly observed sales in a number of stores for one or more chains in a certain region. The sales data can concern any available product category, although usually one keeps track of products that are not perishable, or at least not immediately. Typical sample sizes range from 2 to 8 years. Usually, one also collects information on marketing instruments as "display", "feature", and "price". Preferably, the price variable can be decomposed into the regular price and the actual price. As such, one can analyze the effects of changes in the regular price and in price promotions (the actual price relative to the regular price). When one considers product categories, one collects data on all brands and stock keeping units (SKUs). Subsequently, these data can be aggregated concerning large national brands, private label brands and a rest category including all smaller brands. The data are obtained through optic scanners. With these data one can analyze the short-run and long-run effects of, what is called, the marketing-mix (the interplay of price setting, promotions, advertising and so on), and also the reactions to and from competitors. A typical graph of such weekly sales data is given in Figure 1, where a large amount of substantial spikes can be noticed. Obviously, one might expect that these observations correspond with increased marketing efforts, and hence one should not delete these data points.

An important area concerns the (dynamic) effects of advertising (or any other

[3] Of course, as with any set of data in any discipline, marketing data can contain outliers, influential data, missing data, censored data, and so on. This aspect is not considered any further here.

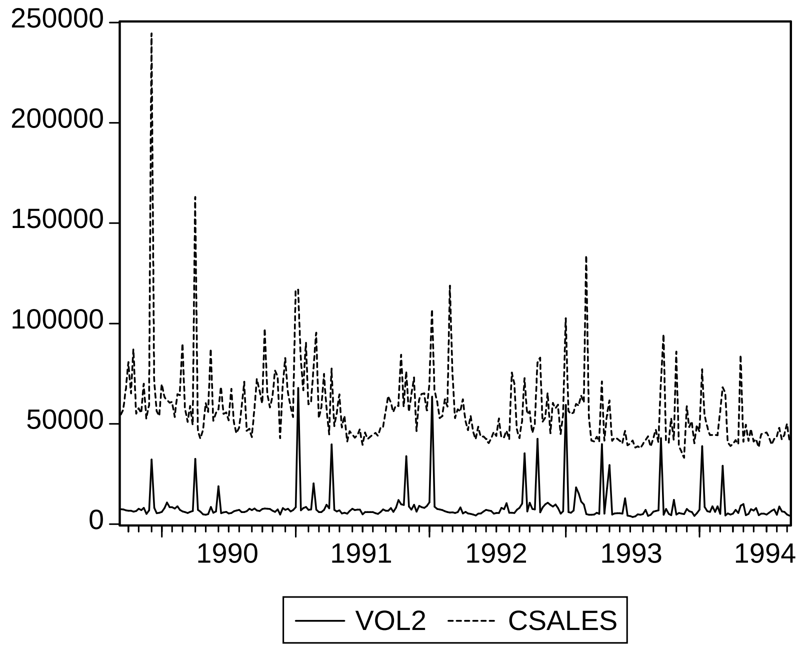


Figure 1: Sales of a brand and category sales, weekly data, 1989-1994, Dominick's Finer Foods

instrument) on sales. How long do these effects last? And, what is most interesting to econometricians, what is the appropriate data interval to estimate these effects? This topic has important implications for marketers, policy makers, and legal scholars. For managers, the duration of the advertising effects has implications for planning and cost allocation. If the effects of advertising last beyond the current period, the true cost of that advertising must be allocated over the relevant time period. And, if the effects of advertising decay slowly and last for decades, advertising may have to be treated as an investment rather than as an expense.

The duration of this so-called advertising carryover can have important legal implications. If the effects of advertising last for decades, firms involved in deceptive advertising would have to be responsible for remedies years and even decades after such a deception occurred. Similarly, firms might be responsible for the advertising they carried out several decades earlier.

The available data on sales and advertising often concern annual or at best monthly data. Unfortunately, for the analysis of short-run and carry-over effects, one may want to have data at a higher frequency. An intriguing data set is presented and analyzed in Tellis, Chandy, and Thaivanich (2000). The advertiser in their study is a medical referral service. The firm advertises a toll free number which customers can call to get the phone number and address of medical service providers. Consumers know the service by the advertised brand name that reflects the toll free number that is advertised. When a customer calls the number, a representative of the firm answers the call. The representative queries the customer and then recommends a suitable service-provider based on location, preferences, and specific type of service needed. Typically, the representative tries to connect the customer to the service-provider directly by phone, again bearing in mind the quoted statement in Kotler *et al.* (2000). Any resulting contact between a customer and the service provider is called a referral. Customers do not pay a fee for the referral, but service providers pay a fixed monthly fee for a specific minimum number of referrals a month. The firm screens service providers before including them as clients. The firm began operations in March 1986 in the Los Angeles market with 18 service providers and a USD 30,000 monthly advertising budget. Around 2000 it advertised in over 62 major markets in the U.S., with a multi-million dollar advertising budget that includes over 3500 TV advertising exposures per month. The primary marketing variable that affects referrals is advertising. A nice aspect of this data set is that it contains observations per hour, and I will return to this particular data set below.

**Market shares**

Market shares are usually defined by own sales divided by category sales. There are various ways to do calculate market shares, where choices have to be made concerning how to measure sales and prices. Next, one might weight the sales of competitors depending on the availability across outlets.

One reason to analyze shares instead of sales is that their time series properties can be more easy to exploit for forecasting. Outlying observations in category sales and in own sales might cancel out, at least approximately. The same holds for seasonality, and even perhaps for trends of the unit root type. Indeed, various empirical studies suggest that, at least for mature markets, market shares tend to be stationary, while sales data might not be. A second reason to analyze market shares is that it directly shows how well the own brand or product fares as compared with competitors. Indeed, if category sales increase rapidly, and own sales only little, then own market share declines, reflecting the descending power of the brand within the category.

An argument against using market shares is that models for sales allow to include and jointly forecast category sales. Also, the introduction of new brands in the observed sample period is more easy to handle than in market share models[4].

It is important for the material below to reiterate the obvious relation between market shares and sales, as it is a non-linear one. Take $S_t$ as own sales and $CS_t$ as category sales, then market share $M_t$ is defined as

$$M_t = \frac{S_t}{CS_t}. \tag{1}$$

As the right hand side is a ratio, it holds that

$$\mathrm{E}(M_t) \neq \frac{\mathrm{E}(S_t)}{\mathrm{E}(CS_t)}, \tag{2}$$

where E denotes the expectations operator. Additionally, $CS_t$ contains $S_t$, and hence the denominator and the numerator are not independent.

Typical graphs of weekly market shares appear in Figure 2. Again one can infer various spikes in one series, and now also similar sized spikes but with different signs for the competitive brands' market shares.

**New product diffusion**

The data on the adoption of a new product, which usually concerns durable products like computers, refrigerators, and CD-players, typically show a sigmoid shape. Often the data concern only annual data for 10 to 20 years. See for example the data depicted in Figure 3, which concern the fraction of music recordings that are sold on compact discs, see Bewley and Griffiths (2003). This sigmoid pattern reflects a typical product life cycle, which starts with early innovators to purchase the product, and which ends with laggards who purchase a new product once almost everyone else already has it.

[4] Fok and Franses (2004) provide a solution for the latter situation.

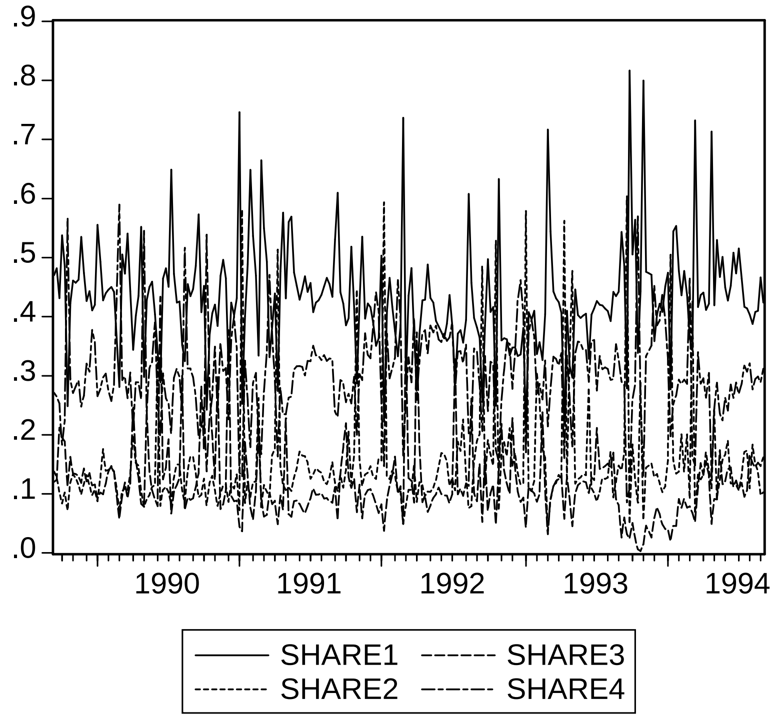


Figure 2: Market shares for four brands of crackers (one is "rest"), weekly data, 1989-1994, Dominick's Finer Foods

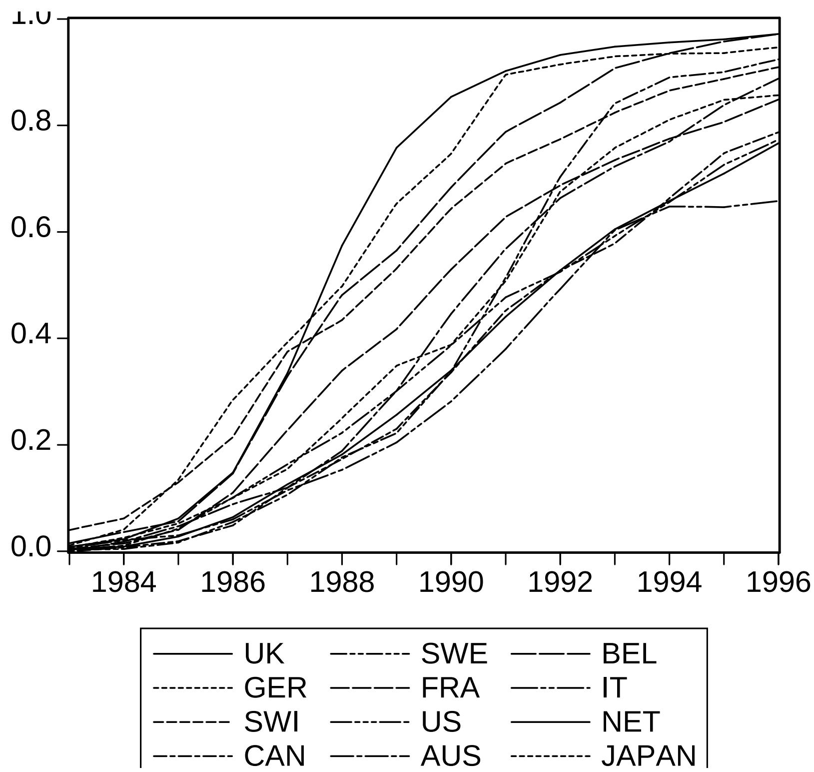


Figure 3: Market penetration of compact discs, 1983-1996

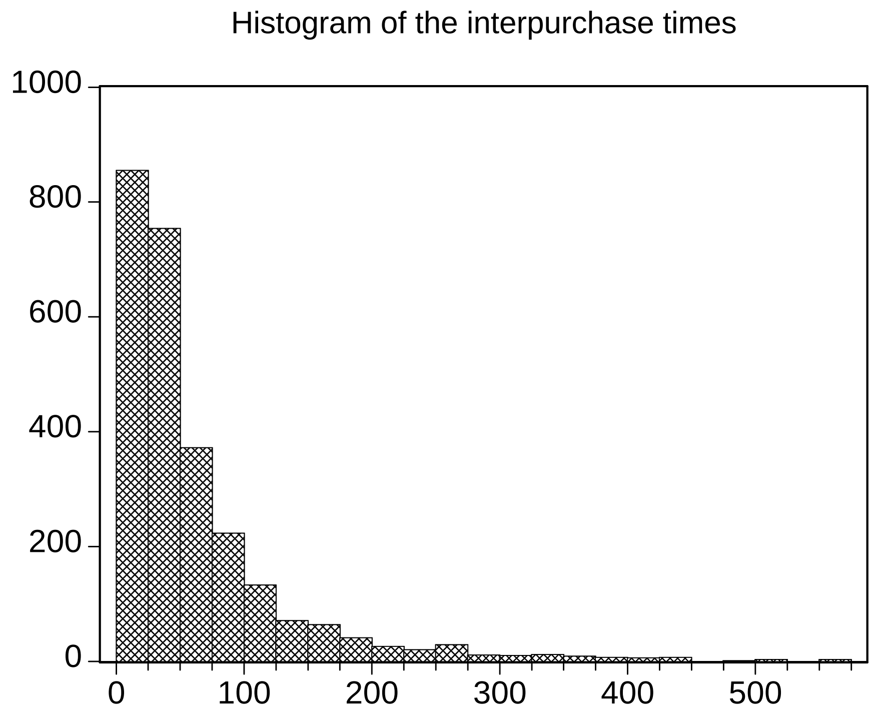


Figure 4: Histogram of the number of days between two liquid detergent purchases

If the diffusion process is measured in terms of fractions of households owning a product, such data are bounded from below and from above. Hence, the model to be used shall somehow need to impose restrictions also as the data span usually is rather short and as one tends to want to make forecasts closer towards the beginning of the diffusion process than towards the end.

**Panels with $N$ and $T$ both large**

Finally, various data in marketing are obtained from observing a sample of $N$ households over $T$ periods. There are household panels with size $N$ around 5000. These keep track of what these households purchase as the households have optic scanners at home, which they use again to document what they had bought on their latest shopping trip. This way one can get information on the choice that individuals make, whether they respond to promotions, and their time between purchases. A typical graph of such interpurchase time appears in Figure 4. In fact, such data allow for a full description of consumption behavior, see for example van Oest *et al.* (2002).

Retail stores keep track of the behavior of their loyalty program members and keep track of everything they purchase (and not purchase). Charities store past donation data of millions of their donators, and insurance firms keep track of all contacts they have with their clients. Those contacts can be telephone calls made by the client to ask for information, but can also be direct mailings sent to them. Sometimes these data are censored or truncated, like in the case of a charity's direct mailing where only those who received a mailing can decide to donate a certain amount or not to donate.

## 2.2 What does one want to forecast?

Usually, these performance measures are of focal interest in a forecasting exercise. Depending on the data and on the question at hand, this can be done either for new cross sections or for future time series data. For example, for new product diffusion it is of interest to forecast whether a product that was recently launched in country A, will also fly in country B. Another example concerns a new list of addresses of potential donators to charity, which cannot all be mailed and a selection will have to made. One then looks for those individuals who are most likely to donate, where these individuals are somehow matched with similar individuals whose track record is already in the database and who usually donate. Additionally, one wants to forecast the effects of changes in marketing instruments like price and promotion on own future sales and own market shares.

In at least two situations forecasting in marketing concerns a little less straightforward situation. The first concerns sales and market shares. The reason is that one usually not only wants to forecast own sales and category sales, but preferably also the response of all competitors to own marketing efforts. This entails that econometric models will contain multiple equations, even in case the interest only lies in own market shares.

A second typical forecasting situation concerns the adoption process of a new product. Usually one wants to make a forecast of the pattern of new to launch products, based on the patterns of related products that have already been introduced. This should also deliver a first guess value of the total amount of adoptions at the end of the process. For that matter, one needs a certain stylized functional form to describe a typical adoption process, with parameters that can be imposed onto the new situation. Moreover, once the new product is brought to the market, one intends to forecast the "take-off" point (where the increase in sales is fastest) and the inflection point (where the level of the sales is highest). As will be seen in the next section, a commonly used model for this purpose is a model with just three parameters, where these parameters directly determine these important change points in the process.

# 3 Models typical to marketing

The type of data and the research question guide the choice of the econometric model to be used. In various situations in marketing research, one can use the standard regression model or any of its well-known extensions. Also, one sees a regular use of the logit or probit model for binomial data, and of the ordered regression model for ordered data, and of the multinomial logit or probit model for unordered multinomial data. Interestingly, the use of the, not that easy to analyze, multinomial probit model is often seen, and this perhaps due to the assumption of the independence of irrelevant alternatives is difficult to maintain in brand choice analysis. Furthermore, one sees models for censored and truncated data, and models for duration and count data. Franses and Paap (2001) provide a summary of the most often used models in marketing research. However, they do not address in detail the econometric models that are specifically found in marketing, and less so elsewhere. This is what I will do in this chapter. These models are the Koyck model to relate advertising with sales, the attraction model to describe market shares, the Bass model for the adoption of new products, and the multi-level regression model for panels of time series. Each of these four types of models will be discussed in the next four subsections.

## 3.1 Dynamic effects of advertising

An important measure to understand the dynamic effects of advertising, that is, how long do advertising pulses last, is the so-called $p$-percent duration interval, see

Clark (1976), Tellis (1988), and Leone (1995), among others. A $p$-percent duration interval measures the time lag between an advertising impulse and the moment that $p$ percent of its effect on sales has decayed.

Denote $S_t$ as sales and $A_t$ as advertising, and assume for the moment that there are no other marketing activities and no competitors. A reasonable model to start with would be an autoregressive distributed lags model of order (p,m) (ADL(p,m)). This model is written as

$$S_t = \mu + \alpha_1 S_{t-1} + \ldots + \alpha_p S_{t-p} + \beta_0 A_t + \beta_1 A_{t-1} + \ldots + \beta_m A_{t-m} + \varepsilon_t. \tag{3}$$

This model implies that

$$\begin{aligned}
\frac{\partial S_t}{\partial A_t} &= \beta_0 \\
\frac{\partial S_{t+1}}{\partial A_t} &= \beta_1 + \alpha_1 \frac{\partial S_t}{\partial A_t} \\
\frac{\partial S_{t+2}}{\partial A_t} &= \beta_2 + \alpha_1 \frac{\partial S_{t+1}}{\partial A_t} + \alpha_2 \frac{\partial S_t}{\partial A_t} \\
&\vdots \\
\frac{\partial S_{t+k}}{\partial A_t} &= \beta_k + \sum_{j=1}^{k} \alpha_j \frac{\partial S_{t+(k-j)}}{\partial A_t}
\end{aligned}$$

where $\alpha_k = 0$ for $k > p$, and $\beta_k = 0$ for $k > m$. These partial derivatives can be used to compute the decay factor

$$p(k) = \frac{\frac{\partial S_t}{\partial A_t} - \frac{\partial S_{t+k}}{\partial A_t}}{\frac{\partial S_t}{\partial A_t}} \tag{4}$$

Due to the very nature of the data, this decay factor can only be computed for discrete values of $k$. Obviously, this decay factor is a function of the model parameters. Through interpolation one can decide on the value of $k$ for which the decay factor is equal to some value of $p$, which is typically set equal to 0.95 or 0.90. This estimated $k$ is then called the $p$-percent duration interval.

Next to its point estimate, one would also want to estimate the confidence bounds of this duration interval, taking aboard that the decay factors are based on non-linear functions of the parameters. The problem when determining the expected value of $p(k)$ is that the expectation of this non-linear function of parameters is not equal to the function applied to the expectation of the parameters, that is $\text{E}(f(\theta)) \neq f(\text{E}(\theta))$. So, the values of $p(k)$ need to be simulated. With the proper assumptions, for the general ADL model it holds that the OLS estimator is asymptotically normal distributed. Franses and Vroomen (2003) suggest to use a large number of simulated parameter vectors from this multivariate normal distribution, and calculate the values of $p(k)$. This simulation exercise also gives the relevant confidence bounds.

**The Koyck model**

Although the general ADL model seems to gain popularity in advertising-sales modeling, see Tellis *et al.* (2000) and Chandy *et al.* (2001), a commonly used model still is the so-called Koyck model. Indeed, matters become much more easy for the ADL model if it is assumed that $m$ is $\infty$, all $\alpha$ parameters are zero and additionally that $\beta_j = \beta_0\lambda^{j-1}$, where $\lambda$ is assumed to be in between 0 and 1. As this model involves an infinite number of lagged variables, one often considers the so-called Koyck transformation (Koyck, 1954). In many studies the resultant model is called the Koyck model[5].

The Koyck transformation amounts to multiplying both sides of

$$S_t = \mu + \beta_0 A_t + \beta_0\lambda A_{t-1} + \beta_0\lambda^2 A_{t-2} + \ldots + \beta_\infty\lambda^\infty A_{t-\infty} + \varepsilon_t \tag{5}$$

with $(1 - \lambda L)$, where $L$ is the familiar lag operator, to get

$$S_t = \mu^* + \lambda S_{t-1} + \beta_0 A_t + \varepsilon_t - \lambda\varepsilon_{t-1}. \tag{6}$$

The short-run effect of advertising is $\beta_0$ and the long-run or total effect is $\frac{\beta_0}{1-\lambda}$. As $0 < \lambda < 1$, the Koyck model implies that the long-run effect exceeds the short-run effect. The $p$-percent duration interval for this model has a convenient explicit expression and it is equal to $\frac{\log(1-p)}{\log\lambda}$.

Even after 50 years, the Koyck model is often used and still stimulates new research, see Franses (2004). For example, the Koyck model involves the familiar Davies (1987) problem. That is, under the null hypothesis that $\beta_0 = 0$, the model

$$S_t = \mu^* + \lambda S_{t-1} + \beta_0 A_t + \varepsilon_t - \lambda\varepsilon_{t-1}, \tag{7}$$

[5] Leendert Marinus Koyck (1918-1962) was a Dutch economist who studied and worked at the Netherlands School of Economics, which is now called the Erasmus University Rotterdam.

collapses into

$$S_t = \mu^* + \varepsilon_t, \tag{8}$$

where $\lambda$ has disappeared. Solutions based on the suggestions in Andrews and Ploberger (1994) and Hansen (1996) are proposed in Franses and Van Oest (2004), where also the relevant critical values are tabulated.

**Temporal aggregation and the Koyck model**

Temporal aggregation entails that one has to analyze data at a macro level while the supposedly true link between sales and advertising happens at a higher frequency micro level. This is particularly relevant nowadays, where television commercials last for just 30 seconds, while sales data are available perhaps only at the daily level. There has been substantial interest in handling the consequences of temporal aggregation in the marketing literature, see Bass and Leone (1983), Assmus *et al.* (1984), Clarke (1976), Leone (1995) and Russell (1988). These studies all impose strong assumptions about the advertising process. A common property of all studies is that they warn about using the same model for micro data and for macro data, as in that case the duration interval will be overestimated, when relying on macro data only.

Recently, Tellis and Franses (2004) argue that only a single assumption is needed for the Koyck model parameters at the micro frequency to be retrievable from the available macro data. This assumption is that the macro data are $K$-period sampled micro data and that there is only a single advertising pulse at time $i$ within that $K-$period. The size of the pulse is not relevant nor is it necessary to know the dynamic properties of the advertising process. This is because this particular assumption for advertising entails that the $K-$period aggregated pulse data match with the size of the single pulse within that period.

Consider again the $K-$period data, and assume that the pulse each time happens at time $i$, where $i$ can be 1, 2, or, $K$. It depends on the location of $i$ within the $K$ periods whether the pulse will be assigned to $A_T$ or $A_{T-1}$, where capital $T$ indicates the macro data. Along these lines, Tellis and Franses (2004) show that the Koyck model for the micro data leads to the following extended Koyck model for $K$-period aggregated data, that is,

$$S_T = \lambda^K S_{T-1} + \beta_1 A_T + \beta_2 A_{T-1} + \varepsilon_T - \lambda^K \varepsilon_{T-1}, \tag{9}$$

with

$$\beta_1 = \beta_0(1 + \lambda + ... + \lambda^{K-i}), \tag{10}$$

and

$$\beta_2 = \beta_0(\lambda^{K-i+1} + ... + \lambda^{K-1}), \tag{11}$$

and where $\beta_2 = 0$ if $i = 1$.

As the parameters for $S_{T-1}$ and $\varepsilon_{T-1}$ are the same, Franses and van Oest (2004) recommend to use estimation by maximum likelihood. The total effect of advertising, according to this extended Koyck model for $K-$period aggregated data, is equal to

$$\frac{\beta_1 + \beta_2}{1 - \lambda^K} = \frac{\beta_0(1 + \lambda + ... + \lambda^{K-i}) + \beta_0(\lambda^{K-i+1} + ... + \lambda^{K-1})}{1 - \lambda^K} = \frac{\beta_0}{1 - \lambda}. \tag{12}$$

Hence, one can use this extended model for the aggregated data to estimate the long-run effects at the micro frequency. Obviously, $\lambda$ can be estimated from $\lambda^K$, and therefore one can also retrieve $\beta_0$.

To illustrate, consider the Miami market with 10776 hourly data, as discussed in Tellis, Chandy and Thaivanich (2000). Given the nature of the advertising data, it seems safe to assume that the micro frequency is 30 seconds. Unfortunately, there are no sales or referrals data at this frequency. As the hour is the least integer time between the exposures, $K$ might be equal to 120, as there are 120 times 30 seconds within an hour. As the advertising pulse usually occurs right after the entire hour, it is likely that $i$ is close to or equal to $K$. The first model I consider is the extended Koyck model as in (9) for the hourly data. I compute the current effect, the carry-over effect and the 95 percent duration interval. Next, I estimate an extended Koyck model for the data when they are aggregated up to days. In this case daily dummy variables are included to capture seasonality to make sure the model fits adequately to the data. The estimation results are summarized in Table 1.

Table 1 shows that the 95 percent duration interval at the 30 seconds frequency is 1392.8. This is equivalent with about 11.6 hours, which is about half a day. In

Table 1: Estimation results for extended Koyck models for hourly and daily data.

| Parameter | Hourly frequency[1] | Daily frequency[2] |
|---|---|---|
| Current effect ($\beta_0$) | 0.008648 | 1.4808 |
| Carry-over effect ($\frac{\beta_0}{1-\lambda}$) | 4.0242 | 5.2455 |
| 95 per cent duration interval | 1392.8 (30 seconds) | 218.77 (days) |

[1] The model estimated for the hourly frequency assumes that the micro frequency is 30 seconds, and that the aggregation level is 120, amounting to hours. The $\lambda$ parameter is estimated to be equal to 0.997851, as $\hat{\lambda}^K$ is 0.772504. There are 10776 hourly observations. The parameter $\beta_2$ is not significant, which suggests that $i$ is indeed close to or equal to $K$.

[2] The model for the 449 daily data is again the extended Koyck model, which includes current and lagged advertising. The model also includes 6 daily dummy variables to capture deterministic seasonality. The $\lambda$ parameter is estimated to be equal to 0.9864.

sharp contrast, if I consider the Koyck model for daily data, I find that this duration interval is about 220 days, or about 7 months. This shows that using the same model for different frequencies can lead to serious overestimation of the duration interval. Of course, the proper model in this case is the extended Koyck model at the hourly frequency, which takes into account that the micro frequency is 30 seconds.

## 3.2 The attraction model for market shares

A market share attraction model is a useful tool for analyzing competitive structure across, for example, brands within a product category. The model can be used to infer cross-effects of marketing-mix variables, but one can also learn about the effects of own efforts while conditioning on competitive reactions. Various details can be found in Cooper and Nakanishi (1988) and various econometric aspects are given in Fok *et al.* (2002).

Important features of an attraction model are that it incorporates that market shares sum to unity and that the market shares of all individual brands are in between 0 and 1. Hence, also forecasts are restricted to be in between 0 and 1. The model (which bears various resemblances with the multinomial logit model) consists of two components. There is a specification of the attractiveness of a brand and a definition of market shares in terms of this attractiveness.

First, define $A_{i,t}$ as the attraction of brand $i$, $i = 1, \ldots, I$ at time $t$, $t = 1, \ldots, T$. This attraction is assumed to be an unobserved (latent) variable. Commonly, it assumed that this attraction can be described by

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) \prod_{j=1}^{I} \prod_{k=1}^{K} x_{k,j,t}^{\beta_{k,j,i}} \tag{13}$$

where $x_{k,j,t}$ denotes the $k$-th explanatory variable (such as price level, distribution, advertising spending) for brand $j$ at time $t$ and where $\beta_{k,j,i}$ is the corresponding coefficient for brand $i$. The parameter $\mu_i$ is a brand-specific constant. Let the error term $(\varepsilon_{1,t}, \ldots, \varepsilon_{I,t})'$ be normally distributed with zero mean and $\Sigma$ can be non-diagonal. Note that data availability determines how many parameters can be estimated in the end, as in this representation (13) there are $I + I + I \times I \times K = I(2 + IK)$ parameters. The $x_{k,j,t}$ is assumed to be non-negative, and hence rates of change are usually not allowed. The variable $x_{k,j,t}$ may be a 0/1 dummy variable to indicate the occurrence of promotional activities for brand $j$ at time $t$. Note that in this case one should transform $x_{k,j,t}$ to $\exp(x_{k,j,t})$ to avoid that attraction becomes zero in case of no promotional activity.

The fact that the attractions are not observed makes the inclusion of dynamic structures a bit complicated. For example for the model

$$A_{i,t} = \exp(\mu_i + \varepsilon_{i,t}) A_{i,t-1}^{\gamma_i} \prod_{j=1}^{I} \prod_{k=1}^{K} x_{k,j,t}^{\beta_{k,j,i}} \tag{14}$$

one can only retrieve $\gamma_i$ if it is assumed that $\gamma = \gamma_i$ for all $i$. Fok *et al.* (2002) provide a detailed discussion on how to introduce dynamics into attraction models.

The second component of the model is simply

$$M_{i,t} = \frac{A_{i,t}}{\sum_{j=1}^{I} A_{j,t}}, \tag{15}$$

which states that market share is the own attraction divided by total attraction. These two equations complete the attraction model.

To enable parameter estimation, one simply takes one of the brands as the benchmark, say, brand $I$. Next, one divides both sides of (15) by $M_{I,t}$, takes natural logarithms of both sides to arrive at a $(I-1)$-dimensional set of equations given by

$$\log M_{i,t} - \log M_{I,t} = (\mu_i - \mu_I) + \sum_{j=1}^{I} \sum_{k=1}^{K} (\beta_{k,j,i} - \beta_{k,j,I}) \log x_{k,j,t} + \eta_{i,t} \tag{16}$$

for $i = 1, \ldots, I-1$. Note that the $\mu_i$ parameters ($i = 1, \ldots, I$) are not identified. In fact, only the parameters $\tilde{\mu}_i = \mu_i - \mu_I$, and $\tilde{\beta}_{k,j,i} = \beta_{k,j,i} - \beta_{k,j,I}$ are identified. This is not problematic for interpretation as the instantaneous elasticity of the $k$-th marketing instrument of brand $j$ on the market share of brand $i$ is given by

$$\begin{aligned} \frac{\partial M_{i,t}}{\partial x_{k,j,t}} \frac{x_{k,j,t}}{M_{i,t}} &= \beta_{k,i,j} - \sum_{r=1}^{I} M_{r,t} \beta_{k,r,j} && (17) \\ &= (\beta_{k,j,i} - \beta_{k,j,I})(1 - M_{i,t}) - \sum_{r=1 \wedge r \neq i}^{I-1} M_{r,t} (\beta_{k,j,r} - \beta_{k,j,I}). && (18) \end{aligned}$$

The attraction model has often been applied in marketing, see Leeflang and Reuyl (1984), Naert and Weverbergh (1981), Kumar (1994), Klapper and Herwartz (2000) and several recent studies. Usually, the model is used for out-of-sample forecasting and to evaluate competitive response, see Bronnenberg, Mahajan and Vanhonacker (2000). Fok and Franses (2004) introduce a version of the model that can be used to describe the consequences of a new entrant in the product category.

Despite the fact that the model is often used for forecasting, the proper way to generate forecasts is not trivial, and in fact, rarely considered in detail. The reason for this non-triviality is that the set of seemingly unrelated regression equations is formulated in terms of the logs of ratios of market shares. However, in the end one intends to forecast the market shares themselves. In the next section, I will demonstrate how appropriate forecasts can be generated.

## 3.3 The Bass model for adoptions of new products

The diffusion pattern of adoptions of new products shows a typical sigmoid shape. There are many functions that can describe such a shape, like the logistic function or the Gompertz function. In marketing research, one tends to focus on one particular function, which is the one proposed in Bass (1969). Important reasons for this are that the model captures a wide range of possible shapes (for example, the logistic function assumes symmetry around the inflection point while the Bass model does not) and that the model parameters can be assigned a workable interpretation.

The Bass (1969) theory starts with a population of $m$ potential adopters. For each of these, the time to adoption is a random variable with a distribution function $F(\tau)$ and density $f(\tau)$, and a hazard rate assumed to be

$$\frac{f(\tau)}{1 - F(\tau)} = p + qF(\tau), \tag{19}$$

where $\tau$ refers to continuous time. The parameters $p$ and $q$ are associated with innovation and imitation, respectively. In words, this model says that the probability of adoption at time $t$, given that no adoption has occurred yet, depends on a constant $p$, which is independent of any factor, hence innovation, and on a fraction of the cumulative density of adoption, hence imitation.

The cumulative number of adopters at time $\tau$, $N(\tau)$, is a random variable with mean $\bar{N}(\tau) = E[N(\tau)] = mF(\tau)$. The function $\bar{N}(\tau)$ satisfies the differential equation

$$\bar{n}(\tau) = \frac{d\bar{N}(\tau)}{d\tau} = p[m - \bar{N}(\tau)] + \frac{q}{m}\bar{N}(\tau)[m - \bar{N}(\tau)]. \tag{20}$$

The solution of this differential equation for cumulative adoption is

$$\bar{N}(\tau) = mF(\tau) = m\left[\frac{1 - e^{-(p+q)\tau}}{1 + \frac{q}{p}e^{-(p+q)\tau}}\right], \tag{21}$$

and for adoption itself it is

$$\bar{n}(\tau) = mf(\tau) = m\left[\frac{p(p+q)^2 e^{-(p+q)\tau}}{\left(p + qe^{-(p+q)\tau}\right)^2}\right], \tag{22}$$

see Bass (1969) for details. Analyzing these two functions of $\tau$ in more detail reveals that $\bar{N}(\tau)$ indeed has a sigmoid pattern, while $\bar{n}(\tau)$ is hump-shaped. Note that the parameters $p$ and $q$ exercise a non-linear impact on the pattern of $\bar{N}(t)$ and $\bar{n}(t)$. For example, the inflection point $T^*$, which corresponds with the time of peak adoptions, equals

$$T^* = \frac{1}{p+q}\log(\frac{q}{p}). \tag{23}$$

Substituting this expression in (21) and in (22), allows a determination of the amount of sales at the peak as well as the amount of the cumulative adoptions at that time.

In practice one of course only has discretely observed data. Denote $X_t$ as the adoptions and $N_t$ as the cumulative adoptions, where $t$ often refers to months or years. There are now various ways to translate the continuous time theory to models for the data on $X_t$ and $N_t$. Bass (1969) proposes to consider the regression model

$$\begin{aligned} X_t &= p(m - N_{t-1}) + \frac{q}{m} N_{t-1}(m - N_{t-1}) + \varepsilon_t \\ &= \alpha_1 + \alpha_2 N_{t-1} + \alpha_3 N_{t-1}^2 + \varepsilon_t, \end{aligned} \tag{24}$$

where it is assumed that $\varepsilon_t$ is an independent and identically distributed error term with mean zero and common variance $\sigma^2$. Note that $(p, q, m)$ must be obtained from $(\alpha_1, \alpha_2, \alpha_3)$, but that for out-of-sample forecasting one can use (24), and hence rely on ordinary least squares (OLS).

Recently, Boswijk and Franses (2005) extend this basic Bass regression model by allowing for heteroskedastic errors and by allowing for short-run deviations from the deterministic S-shaped growth path of the diffusion process, as implied by the differential equation in (20). The reason to include heteroskedasticity is that, in the beginning and towards the end of the adoption process, one should be less uncertain about the variance of the forecasts than when the process is closer to the inflection point. Next, the solution to the differential equation is a deterministic path, and there may be various reasons to temporally deviate form this path. Boswijk and Franses (2005) therefore propose to consider

$$dn(\tau) = \alpha \left[ p[m - N(\tau)] + \frac{q}{m} N(\tau)[m - N(\tau)] - n(\tau) \right] d\tau + \sigma n(\tau)^\gamma dW(\tau), \tag{25}$$

where $W(\tau)$ is a standard Wiener process. The parameter $\alpha$ in (25) measures the speed of adjustment towards the deterministic path implied by the standard Bass model. Additionally, by introducing $\sigma n(t)^\gamma$, heteroskedasticity is allowed. A possible choice is to set $\gamma = 1$. Boswijk and Franses (2005) further derive that the discretization of this continuous time model is

$$X_t - X_{t-1} = \beta_1 + \beta_2 N_{t-1} + \beta_3 N_{t-1}^2 + \beta_4 X_{t-1} + X_{t-1}\varepsilon_t, \tag{26}$$

where

$$\beta_1 = \alpha p m \tag{27}$$

$$\beta_2 = \alpha(q-p) \tag{28}$$

$$\beta_3 = -\alpha\frac{q}{m} \tag{29}$$

$$\beta_4 = -\alpha, \tag{30}$$

which shows that all parameters in (26) depend on $\alpha$.

Another empirical version of the Bass theory, a version which is often used in practice, is proposed in Srinivasan and Mason (1986). These authors recognize that the Bass (1969) formulation above may introduce aggregation bias, as $X_t$ is simply taken as the discrete representative of $n(\tau)$. Therefore, Srinivasan and Mason (1986) propose to apply non-linear least-squares (NLS) to

$$X_t = m[F(t;\theta) - F(t-1;\theta)] + \varepsilon_t, \tag{31}$$

where $\theta$ collects $p$ and $q$. Van den Bulte and Lilien (1997) show that this method is rather unstable if one has data that do not yet cover the inflection point. How to derive forecasts for the various models will be discussed below.

## 3.4 Multi-level models for panels of time series

It is not uncommon in marketing to have data on a large number of cases (households, brands, SKUs) for a large number of time intervals (like a couple of years with weekly data). In other words, it is not uncommon that one designs models for a variable to be explained with substantial information over dimension $N$ as well as $T$. Such data are called a panel of time series. Hence, one wants to exploit the time series dimension, and potentially include seasonality and trends, while preserving the panel structure.

To set notation, consider

$$y_{i,t} = \mu_i + \beta_i x_{i,t} + \varepsilon_{i,t}, \tag{32}$$

where subscript $i$ refers to household $i$ and $t$ to week $t$. Let $y$ denote sales of a certain product and $x$ be price, as observed by that particular household (where a household can visit a large variety of stores).

### Hierarchical Bayes approach

It is not uncommon to allow the $N$ households to have different price elasticities. And, from a statistical perspective, if one were to impose $\beta_i = \beta$, one for sure would reject this hypothesis in most practical situations. On the other hand, the interpretation of $N$ different price elasticities is also not easy either. Typically, one does have a bit more information on the households (family life cycle, size, income, education), and it might be that these variables have some explanatory value for the price elasticities. One way to examine this would be to perform $N$ regressions, to retrieve the $\hat{\beta}_i$, and next, in a second round, to regress these estimated values on household-specific features. Obviously, this two-step approach assumes that the $\hat{\beta}_i$ variables are given instead of estimated, and hence, uncertainty in the second step is underestimated.

A more elegant solution is to add a second level to (32), that is for example

$$\beta_i \sim \mathrm{N}(\beta_0 + \beta_1 z_i, \sigma^2), \tag{33}$$

where $z_i$ is an observed variable for a household, see Blattberg and George (1991). Estimation of the model parameters can require simulation-based techniques. An often used method is termed Hierarchical Bayes (HB), see Allenby and Rossi (1999) among various others.

An exemplary illustration of this method given in Van Nierop, Fok and Franses (2002) who consider this model for 2 years of weekly sales on 23 items in the same product category. The effects of promotions and distribution in $x_{i,t}$ are made a function of the size of an item and its location on a shelf.

### Latent class modeling

As segmentation is often viewed as an important reason to construct models in marketing, another popular approach is to consider the panel model

$$y_{i,t} = \mu_i + \beta_{i,s} x_{i,t} + \varepsilon_{i,t}, \tag{34}$$

where $\beta_{i,s}$ denotes that, say, household-specific price elasticity, can be classified into $J$ classes, within which the price elasticities obey $\beta_{i,s} = \beta(S_i)$, where $S_i$ is element of 1,2,...,$J$, with probability $Pr(S_i = j) = p_j$. In words, $\beta_{i,s}$ corresponds with observation $i$ in class $j$, with $j = 1, 2, ..., J$. Each household has a probability $p_j$, with $p_1 + p_2 + ... + p_J = 1$, to get assigned to a class $j$, at least according to the values of $\beta_{i,s}$. Such a model can be extended to allow the probabilities to depend on household-specific features. This builds on the latent class methodology, recently summarized in Wedel and Kamakura (1999). As such, the model allows for capturing unobserved heterogeneity.

This approach as well as the previous one involves the application of simulation methods to estimate parameters. As simulations are used, the computation of forecasts is trivial. They immediately come as a by-product of the estimation results. Uncertainty around these forecasts can also easily be simulated.

### Panels of time series in other areas

These two classes of models have recently found their way to other disciplines, like macroeconomics. Fok, Franses and van Dijk (2005) introduce a multi-level smooth transition model for a panel of time series, which can be used to examine the presence of common non-linear business cycle features across many variables. The model is positioned in between a fully pooled model, which imposes such common features, and a fully heterogeneous model, which allows for unrestricted non-linearity. They introduce a second-stage model linking the parameters that determine the timing of the switches between business cycle regimes to observable explanatory variables, thereby allowing for different lead-lag relationships across panel members. The model is successfully illustrated using quarterly industrial production in 19 US manufacturing sectors.

Paap, Franses and van Dijk (2005) address the question whether countries on the sub-Saharan African continent have lower average growth rates in real GDP per capita than countries in Asia, Latin and Middle America and the Middle East. In contrast to previous studies, they do not *a priori* assign countries to clusters based on geographical location or other characteristics. Instead, they propose a so-called latent class panel time series model, which allows a data-based classification of countries into clusters such that within a cluster, countries have the same average growth rate. The empirical results suggest that three clusters are sufficient to describe the different growth paths of the countries involved. 26 African countries can be assigned to the low growth cluster, but 8 African countries show growth rates which are comparable with many countries in Asia, Latin and Middle America and in the Middle East.

### A multi-level Bass model

This section is concluded with a brief discussion of a Bass type model for a panel of time series. Talukdar *et al.* (2002) introduce a two-level panel model for a set of diffusion data, where they correlate individual Bass model parameters with explanatory variables in the second stage.

Following the Boswijk and Franses (2005) specification, a panel Bass model would be

$$X_{i,t} - X_{i,t-1} = \beta_{1,i} + \beta_{2,i} N_{i,t-1} + \beta_{3,i} N_{i,t-1}^2 + \beta_{4,i} X_{i,t-1} + X_{i,t-1}\varepsilon_{i,t}. \tag{35}$$

As before, the $\beta$ parameters are functions of the underlying characteristics of the diffusion process, that is,

$$\begin{aligned} \beta_{1,i} &= \alpha_i p_i m_i, & (36) \\ \beta_{2,i} &= \alpha_i (q_i - p_i) & (37) \\ \beta_{3,i} &= -\alpha_i \frac{q_i}{m_i}, & (38) \\ \beta_{4,i} &= -\alpha_i. & (39) \end{aligned}$$

As the effects of $p$ and $q$ on the diffusion patterns are highly non-linear, it seems more appropriate to focus on the inflection point, that is, the timing of peak adoptions, $T_i^*$, and the level of the cumulative adoptions at the peak divided by $m_i$, denoted as $f_i$. The link between $p_i$ and $q_i$ and the inflection point parameters is given by

$$\begin{aligned} p_i &= (2f_i - 1)\frac{\log(1-2f_i)}{2T_i^*(1-f_i)} & (40) \\ q_i &= -\frac{\log(1-2f_i)}{2T_i^*(1-f_i)}, & (41) \end{aligned}$$

see Franses (2003a).

Fok and Franses (2005) propose to specify $\beta_{1,i}, \ldots, \beta_{4,i}$ as a function of the total number of adoptions ($m_i$), the fraction of cumulative adoptions at the inflection point ($f_i$), the time of the inflection point ($T_i^*$), and the speed of adjustment ($\alpha_i$) of $X_{i,t}$ to the equilibrium path denoted as $\beta_{k,i} = \beta_k(m_i, f_i, T_i^*, \alpha_i)$. The adoptions that these authors study are the citations to articles published in *Econometrica* and in the *Journal of Econometrics*. They relate $m_i, f_i, T_i^*$, and $\alpha_i$ to observable features of the articles. In sum, they consider

$$X_{i,t} - X_{i,t-1} = \beta_1(m_i, f_i, T_i^*, \alpha_i) + \beta_2(m_i, f_i, T_i^*, \alpha_i)N_{i,t-1} + \\ \beta_3(m_i, f_i, T_i^*, \alpha_i)N_{i,t-1}^2 + \beta_4(m_i, f_i, T_i^*, \alpha_i)X_{i,t-1} + X_{i,t-1}\varepsilon_{i,t}, \quad (42)$$

where $\varepsilon_{i,t} \sim N(0, \sigma_i^2)$ with

$$\log(m_i) = Z_i'\theta_1 + \eta_{1,i}, \quad (43)$$

$$\log(\frac{2f_i}{1-2f_i}) = Z_i'\theta_2 + \eta_{2,i}, \quad (44)$$

$$\log(T_i^*) = Z_i'\theta_3 + \eta_{3,i}, \quad (45)$$

$$\alpha_i = Z_i'\theta_4 + \eta_{4,i}, \quad (46)$$

$$\log \sigma_i^2 = Z_i'\theta_5 + \eta_{5,i}, \quad (47)$$

where the $Z_i$ vector contains an intercept and explanatory variables.

This section has reviewed various models that are often applied in marketing, and some of which seem to slowly diffuse into other economics disciplines.

# 4 Deriving forecasts

The previous section indicated that various interesting measures (like duration interval) or models (like the attraction model) in marketing research imply that the variable of interest is a non-linear function of variables and parameters. In many cases there are no closed-form solutions to these expressions, and hence one has to resort to simulation-based techniques. In this section the focus will be on the attraction model and on the Bass model, where the expressions for out-of-sample forecasts will be given. Additionally, there will be a discussion of how one should derive forecasts for market shares when forecasts for sales are available.

## 4.1 Attraction model forecasts

As discussed earlier, the attraction model ensures logical consistency, that is, market shares lie between 0 and 1 and they sum to 1. These restrictions imply that (functions of) model parameters can be estimated from a multivariate reduced-form model with $I-1$ equations. The dependent variable in each of the $I-1$ equations is the natural logarithm of a relative market share, that is, $\log m_{i,t} \equiv \log \frac{M_{i,t}}{M_{I,t}}$, for $i = 1, 2, \ldots, I-1$, where the base brand $I$ can be chosen arbitrarily, as discussed before.

In practice, one is usually interested in predicting $M_{i,t}$ and not in forecasting the logs of the relative market shares. Again, it is important to recognize that, first of all, $\exp(\mathrm{E}[\log m_{i,t}])$ is not equal to $\mathrm{E}[m_{i,t}]$ and that, secondly, $\mathrm{E}[\frac{M_{i,t}}{M_{I,t}}]$ is not equal to $\frac{\mathrm{E}[M_{i,t}]}{\mathrm{E}[M_{I,t}]}$. Therefore, unbiased market share forecasts cannot be directly obtained by these data transformations.

To forecast the market share of brand $i$ at time $t$, one needs to consider the relative market shares

$$m_{j,t} = \frac{M_{j,t}}{M_{I,t}} \quad \text{for } j = 1, 2 \ldots, I, \tag{48}$$

as $m_{1,t}, \ldots, m_{I-1,t}$ form the dependent variables (after log transformation) in the reduced-form model. As $M_{I,t} = 1 - \sum_{j=1}^{I-1} M_{j,t}$, it holds that

$$M_{I,t} = \frac{1}{1 + \sum_{j=1}^{I-1} m_{j,t}} \tag{49}$$

$$M_{i,t} = M_{I,t} m_{i,t} = \frac{m_{i,t}}{1 + \sum_{j=1}^{I-1} m_{j,t}} \tag{50}$$

Note that $m_{I,t} = \frac{M_{I,t}}{M_{I,t}} = 1$ and hence (50) can be summarized as

$$M_{i,t} = \frac{m_{i,t}}{\sum_{j=1}^{I} m_{j,t}}, \tag{51}$$

for $i = 1, 2, \ldots, I$.

Fok, Franses and Paap (2002) propose to simulate the one-step ahead forecasts of the market shares as follows. First draw $\eta_t^{(l)}$ from $\mathrm{N}(0, \tilde{\Sigma})$, then compute

$$m_{i,t}^{(l)} = \exp(\tilde{\mu}_i + \eta_{i,t}^{(l)}) \prod_{j=1}^{I} \left( \prod_{k=1}^{K} x_{k,j,t}^{\tilde{\beta}_{k,j,i}} \right), \tag{52}$$

with $m_{I,t}^{(l)} = 1$ and finally compute

$$M_{i,t}^{(l)} = \frac{m_{i,t}^{(l)}}{\sum_{j=1}^{I} m_{j,t}^{(l)}} \quad \text{for } i = 1, \ldots, I, \tag{53}$$

where $l = 1, \ldots, L$ denotes the simulation iteration. Each vector $(M_{1,t}^{(l)}, \ldots, M_{I,t}^{(l)})'$ generated this way is a draw from the joint distribution of the market shares at time $t$. Using the average over a sufficiently large number of draws one can calculate the expected value of the market shares. This can be modified to allow for parameter uncertainty, see Fok, Franses and Paap (2002). Multi-step ahead forecasts can be generated along similar lines.

## 4.2 Forecasting market shares from models for sales

The previous results assume that one is interested in forecasting market shares based on models for market shares. In practice, it might sometimes be more easy to make models for sales. One might then me tempted to divide the own sales forecast by a forecast for category sales, but this procedure leads to biased forecasts for similar reasons as before. A solution is given in Fok and Franses (2001) and will be discussed next.

An often used model (SCAN*PRO) for sales is

$$\log S_{i,t} = \mu_i + \sum_{j=1}^{I} \sum_{k=1}^{K} \beta_{k,j,i} x_{k,j,t} + \sum_{j=1}^{I} \sum_{p=1}^{P} \alpha_{p,j,i} \log S_{j,t-p} + \varepsilon_{i,t}, \tag{54}$$

with $i = 1, \ldots, I$, where $\varepsilon_t \equiv (\varepsilon_{1,t}, \ldots, \varepsilon_{I,t})' \sim N(0, \Sigma)$ and where $x_{k,j,t}$ denotes the $k$-th explanatory variable (for example, price or advertising) for brand $j$ at time $t$ and where $\beta_{k,j,i}$ is the corresponding coefficient for brand $i$, see Wittink et al. (1988). The market share of brand $i$ at time $t$ can of course be defined as

$$M_{i,t} = \frac{S_{i,t}}{\sum_{j=1}^{I} S_{j,t}}. \tag{55}$$

Forecasts of market shares at time $t+1$ based on information on all explanatory variables up to time $t+1$, denoted by $\Pi_{t+1}$, and information on realizations of the sales up to period $t$, denoted by $\mathcal{S}_t$, should be equal to the expectation of the market shares given the total amount of information available, denoted by $\mathrm{E}[M_{i,t+1}|\Pi_{t+1}, \mathcal{S}_t]$, that is,

$$\mathrm{E}[M_{i,t+1}|\Pi_{t+1}, \mathcal{S}_t] = \mathrm{E}\left[ \frac{S_{i,t+1}}{\sum_{j=1}^{I} S_{j,t+1}} \middle| \Pi_{t+1}, \mathcal{S}_t \right]. \tag{56}$$

Due to non-linearity it is therefore not possible to obtain market shares forecasts directly from sales forecasts. A further complication is that it is also not trivial to obtain a forecast of $S_{i,t+1}$, as the sales model concerns log-transformed variables, and it is well known that $\exp(\mathrm{E}[\log X]) \neq \mathrm{E}[X]$. See also Arino and Franses (2000) and Wierenga and Horvath (2005) for the relevance of this notion when examining multivariate time series models. In particular, Wierenga and Horvath (2005) show how to derive impulse response functions from VAR models for marketing variables, and they demonstrate the empirical relevance of a correct treatment of log-transformed data.

Fok and Franses (2001) provide a simulation-based solution. Naturally, unbiased forecasts of the $I$ market shares should be based on the expected value of the market shares, that is,

$$\begin{aligned} &\mathrm{E}[M_{i,t+1}|\Pi_{t+1}, \mathcal{S}_t] = \\ &\quad \int_0^{+\infty} \cdots \int_0^{+\infty} \frac{s_{i,t+1}}{\sum_{j=1}^{I} s_{j,t+1}} f(s_{1,t+1}, \ldots, s_{I,t+1}|\Pi_{t+1}, \mathcal{S}_t) ds_{1,t+1}, \ldots, ds_{I,t+1}, \end{aligned} \tag{57}$$

where $f(s_{1,t+1}, \ldots, s_{I,t+1}|\Pi_{t+1}, \mathcal{S}_t)$ is a probability density function of the sales conditional on the available information, and $s_{i,t+1}$ denotes a realization of the stochastic process $S_{i,t+1}$. The model defined in the distribution of $S_{t+1}$, given $\Pi_{t+1}$ and $\mathcal{S}_t$, is log-normal, but other functional forms can be considered too. Hence,

$$(\exp(S_{1,t+1}), \ldots, \exp(S_{I,t+1}))' \sim N(Z_{t+1}, \Sigma), \tag{58}$$

where $Z_t = (Z_{1,t}, \ldots, Z_{I,t})'$ is the deterministic part of the model, that is,

$$Z_{i,t} = \mu_i + \sum_{j=1}^{I} \sum_{k=1}^{K} \beta_{k,j,i} x_{k,j,t} + \sum_{j=1}^{I} \sum_{p=1}^{P} \alpha_{p,j,i} \log S_{j,t-p} . \tag{59}$$

The $I$-dimensional integral in (57) is difficult to evaluate analytically. Fok and Franses (2001) therefore outline how to compute the expectations using simulation techniques. In short, using the estimated probability distribution of the sales, realizations of the sales are simulated. Based on each set of these realizations of all brands, the market shares can be calculated. The average over a large number of replications gives the expected value in (57).

Forecasting $h > 1$ steps ahead is slightly more difficult as the values of the lagged sales are no longer known. However, for these lagged sales appropriate simulated values can be used. For example, 2-step ahead forecasts can be calculated by averaging over simulated values $M_{i,t+2}^{(l)}$, based on draws $\varepsilon_{t+2}^{(l)}$ from $N(0, \hat{\Sigma})$ and on draws $S_{i,t+1}^{(l)}$, which are already used for the 1-step ahead forecasts. Notice that the 2-step ahead forecasts do not need more simulation iterations than the one-step ahead forecasts.

An important by-product of the simulation method is that it is now also easy to calculate confidence bounds for the forecasted market shares. Actually, the entire distribution of the market shares can be estimated based on the simulated values. For example, the lower bound of a 95% confidence interval is that value for which it holds that 2.5% of the simulated market shares are smaller. Finally, the lower bound and the upper bound always lie within the [0,1] interval, and this should be the case for market shares indeed.

## 4.3 Bass model forecasts

The Bass model is regularly used for out-of-sample forecasting. One way is to have several years of data on own sales, estimate the model parameters for that particular series, and extrapolate the series into the future. As Van den Bulte and Lilien (1997) demonstrate, this approach is most useful in case the inflection point is within the sample. If not, then one might want to consider imposing the parameters obtained for other markets or situations, and then extrapolate.

The way the forecasts are generated depends on the functional form chosen, that is, how one includes the error term in the model. The Srinivasan and Mason (1986) model seems to imply the most easy to construct forecasts. Suppose one aims to predict $X_{n+h}$, where $n$ is the forecast origin and $h$ is the horizon. Then, given the assumption on the error term, the forecast is

$$\hat{X}_{n+h} = \hat{m}[F(n+h;\hat{\theta}) - F(n-1+h;\hat{\theta})]. \tag{60}$$

When the error term is AR(1), straightforward modifications of this formula should be made. If the error term has an expected value equal to zero, then these forecasts are unbiased, for any $h$.

This is in contrast with the Bass regression model, and also its Boswijk and Franses modification, as these models are intrisically non-linear. For one-step ahead, the true observation at $n+1$ in the Bass scheme is

$$X_{n+1} = \alpha_1 + \alpha_2 N_n + \alpha_3 N_n^2 + \varepsilon_{n+1}. \tag{61}$$

The forecast from origin $n$ equals

$$\hat{X}_{n+1} = \hat{\alpha}_1 + \hat{\alpha}_2 N_n + \hat{\alpha}_3 N_n^2 \tag{62}$$

and the squared forecast error is $\sigma^2$. This forecast is unbiased.

For two steps ahead matters become different. Due to the term $N_n^2$, it can be shown that the expected forecast error is

$$E(\hat{X}_{n+2} - X_{n+2}) = -\alpha_3\sigma^2. \tag{63}$$

It is straightforward to derive that if $h$ is 3 or more, this bias grows exponentially with $h$. Naturally, the size of the bias depends on $\alpha_3$ and $\sigma^2$, which both can be small. As the sign of $\alpha_3$ is always negative, the forecast is upward biased.

Franses (2003b) shows that to obtain unbiased forecasts for the Bass-type regression models for $h = 2, 3...$, one needs to resort to simulation techniques. Consider again the Bass regression, now written as

$$X_t = g(Z_{t-1};\pi) + \varepsilon_t, \tag{64}$$

where $Z_{t-1}$ contains 1, $N_{t-1}$ and $N_{t-1}^2$, and $\pi$ includes $p$, $q$ and $m$. A simulation-based one-step ahead forecast is now given by

$$X_{n+1,i} = g(Z_n;\hat{\pi}) + e_i, \tag{65}$$

where $e_i$ is a random draw from the $N(0, \hat{\sigma}^2)$ distribution. Based on $I$ such draws, an unbiased forecast can be constructed as

$$\hat{X}_{n+1} = \frac{1}{I} \sum_{i=1}^{I} X_{n+1,i}. \tag{66}$$

Again, a convenient by-product of this approach is the full distribution of the forecasts. A two-step simulation-based forecast can be based on the average value of

$$X_{n+2,i} = g(Z_n, X_{n+1,i}; \hat{\pi}) + e_i, \tag{67}$$

again for $I$ draws, and so on.

## 4.4 Forecasting duration data

Finally, there are various studies in marketing that rely on duration models to describe interpurchase times. These data are relevant to managers as one can try to speed up the purchase process by implementing marketing efforts, but also one may forecast the amount of sales to be expected in the next period, due to promotion planning. Interestingly, it is known that many marketing efforts have a dynamic effect that stretches beyond the one-step ahead horizon. For example, it has been widely established that there is a so-called post-promotional dip, meaning that sales tend to collapse the week after a promotion was held, but might regain their original level or preferably a higher level after that week. Hence, managers might want to look beyond the one-step ahead horizon.

In sum, one seems to be more interested in the number of purchases in the next week or next month, than that there is an interest in the time till the next purchase. The modelling approach for the analysis of recurrent events in marketing, like the purchase timing of frequently purchased consumer goods, has, however, mainly aimed at explaining the interpurchase times. The main trend is to apply a Cox (mixed) Proportional Hazard model for the interpurchase times, see Seetharaman and Chintagunta (2003) for a recent overview. In this approach after each purchase the duration is reset to zero. This transformation removes much of the typical behavior of the repeat purchase process in a similar way as first-differencing in times series. Therefore, it induces important limitations to the use of time-varying covariates (and also seasonal effects) and duration dependence in the models.

An alternative is to consider the whole path of the repeat purchase history on the time scale starting at the beginning of the observation window. Bijwaard, Franses and Paap (2003) put forward a statistical model for interpurchase times that takes into account all the current and past information available for all purchases as time continues to run along the calendar timescale. It is based on the Andersen and Gill (1982) approach. It delivers forecasts for the number of purchases in the next period **and** for the timing of the first and consecutive purchases. Purchase occasions are modelled in terms of a counting process, which counts the recurrent purchases for each household as they evolve over time. These authors show that formulating the problem as a counting process has many advantages, both theoretically and empirically.

# 5 Conclusion

This chapter has reviewed various aspects of econometric modeling and forecasting in marketing. In many marketing research studies there are quite a number of observations and typically the data are well measured. Usually there is an interest in modeling and forecasting performance measures such as sales, shares, retention, loyalty, brand choice and the time between events, preferably when these depend partially on marketing-mix instruments like promotions, advertising, and price.

Various marketing models are non-linear models. This is due to specific structures imposed on the models to make them more suitable for their particular purpose, like the Bass model for diffusion and the attraction model for market shares. Other models that are frequently encountered in marketing, and less so in other areas (at least as of yet) concern panels of time series. Interestingly, it seems that new econometric methodology (like the Hierarchical Bayes methods) has been developed and applied in marketing first, and will perhaps be more often used in the future in other areas too.

There are two areas in which more research seems needed. The first is that it is not yet clear how out-of-sample forecasts should be evaluated. Of course, mean squared forecast error type methods are regularly used, but it is doubtful whether these criteria meet the purposes of an econometric model. In fact, if the model concerns the retention of customers, it might be worse to underestimate the probability of leaving than to overestimate that probability. Hence the monetary value, possibly discounted for future events, might be more important.

Second, the way forecasts are implemented into actual marketing strategies is not trivial, see Franses (2005a,b). In marketing one deals with customers and with competitors, and each can form expectations about what you will do. The successfulness of a marketing strategy depends on the accuracy of stake-holders' expectations and their subsequent behavior. For example, to predict whether a newly launched product will be successful might need more complicated econometric models than we have available today.

# 6 References

Allenby, G. and P. Rossi (1999), Marketing models of consumer heterogeneity, *Journal of Econometrics*, 89, 57-78.

Andersen, P. K. and R. D. Gill (1982), Cox's regression model for counting processes: A large sample study, *Annals of Statistics* 10, 1100–1120.

Andrews, D.W.K. and W. Ploberger (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica*, 62, 1383-1414.

Arino, M. and P.H. Franses (2000), Forecasting the levels of vector autoregressive log-transformed time series, *International Journal of Forecasting*, 16, 111-116.

Assmus, G., Farley, J.U. and Lehmann, D. (1984), How advertising affects sales: Meta-analysis of econometric results, *Journal of Marketing Research*, 21, 65-74.

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

Bass, F.M. and Leone, R. (1983), Temporal aggregation, the data interval bias, and empirical estimation of bimonthly relations from annual Data, *Management Science*, 29, 1-11.

Bewley, R. and W.E. Griffiths (2003), The penetration of CDs in the sound recording market: issues in specification, model selection and forecasting, *International Journal of Forecasting*, 19, 111–121.

Blattberg, R.C. and E.I. George (1991), Shrinkage estimation of price and promotional elasticities - Seemingly unrelated equations, *Journal of the American Statistical Association*, 86, 304-315.

Boswijk, H.P. and P.H. Franses (2005), On the econometrics of the Bass diffusion model, *Journal of Business and Economic Statistics*, to appear.

Bronnenberg, B.J., V. Mahajan and W.R. Vanhonacker (1999), The emergence of market structure in new repeat-purchase categories: The interplay of market share and retailer distribution, *Journal of Marketing Research*, 37, 16-31.

Bijwaard, G.E, P.H. Franses and R. Paap (2003), Modeling purchases as repeated events, Econometric Institute Report, 2003-45, Erasmus University Rotterdam, Revision requested by the *Journal of Business and Economic Statistics.*

Chandy R.K., Tellis, G.J., MacInnis, D.J., and P. Thaivanich (2001), What to say when: Advertising appeals in evolving markets , *Journal of Marketing Research*, 38, 399-414.

Clarke D.G. (1976), Econometric measurement of the duration of advertising effect on sales, *Journal of Marketing Research*, 8, 345 - 357.

Cooper, L.G. and M. Nakanishi (1988), *Market Share Analysis: Evaluating Competitive Marketing Effectiveness*, Boston: Kluwer Academic Publishers.

Davies, R.B. (1987), Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika*, 64, 247-254.

Fok, D. and P.H. Franses (2001), Forecasting market shares from models for sales, *International Journal of Forecasting*, 17, 121-128.

Fok, D. and P.H. Franses (2004), Analyzing the effects of a brand introduction on competitive structure using a market share attraction model, *International Journal of Research in Marketing*, 21, 159-177.

Fok, D. and P.H. Franses (2005), Modeling the diffusion of scientific publications, *Journal of Econometrics*, to appear.

Fok, D., P.H. Franses and R. Paap (2002), Econometric analysis of the market share attraction model, Chapter 10 in P.H. Franses and A.L. Montgomery (eds.), *Econometric Models in Marketing*, Amsterdam: Elsevier, 223-256.

Fok, D., D.J.C. van Dijk and P.H., Franses (2005), A multi-level panel smooth transition autoregression for US sectoral production, *Journal of Applied Econometrics*, to appear.

Franses, P.H. (2003a), On the diffusion of scientific publications. The case of Econometrica 1987, *Scientometrics*, 56, 29-42

Franses, P.H. (2003b), On the Bass diffusion theory, empirical models and out-of-sample forecasting, ERIM Report Series Research in Management ERS-2003-34-MKT, Erasmus University Rotterdam

Franses, P.H. (2004), Fifty years since Koyck (1954), *Statistica Neerlandica*, to appear.

Franses, P.H. (2005a), On the use of econometric models for policy simulation in marketing, *Journal of Marketing Research*, 42, to appear.

Franses, P.H. (2005b), Diagnostics, expectation, and endogeneity, *Journal of Marketing Research*, 42, to appear.

Franses, P.H. and R. Paap (2001), *Quantitative Models in Marketing Research*, Cambridge: Cambridge University Press.

Franses P.H. and R.D. van Oest (2004), On the econometrics of the Koyck model, Econometric Institute Report 2004-07, Erasmus University Rotterdam.

Franses, P.H. and B. Vroomen (2003), Estimating Duration Intervals, ERIM Report Series Research in Management, ERS-2003-031-MKT, Erasmus University Rotterdam. Revision requested by the *Journal of Advertising.*

Hansen, B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica*, 64, 413-430.

Klapper, D. and H. Herwartz (2000), Forecasting market share using predicited values of competitor behavior: Further empirical results, *International Journal of Forecasting*, 16, 399–421.

Kotler, Ph., D.C. Jain, and S. Maesincee (2002), *Marketing Moves. A New Approach to Profits, Growth, and Renewal*, Boston: Harvard Business School Press.

Koyck, L.M. (1954), *Distributed Lags and Investment Analysis*, Amsterdam: North-Holland.

Kumar, V. (1994), Forecasting performance of market share models: an assessment, additional insights, and guidelines, *International Journal of Forecasting*, 10, 295-312.

Leeflang, P. S. H. and J. C. Reuyl (1984), On the predictive power of market share attraction model, *Journal of Marketing Research*, 21, 211-215.

Leone, R.P. (1995), Generalizing what is known about temporal aggregation and advertising carryover, *Marketing Science*, 14, G141-G150.

Naert, P. A. and M. Weverbergh (1981), On the prediction power of market share attraction models, *Journal of Marketing Research*, 18, 146–153.

Paap, R., P.H. Franses and D. van Dijk (2005), Does Africa grow slower than Asia, Latin America and the Middle East? Evidence from a new data-based classification method, *Journal of Development Economics*, to appear.

Russell, G.J. (1988), Recovering Measures of Advertising Carryover from Aggregate Data: The Role of the Firm's Decision Behavior, *Marketing Science*, 7, 252-270.

Seetharaman, P. B. and P. K. Chintagunta (2003), The proportional hazard model for purchase timing: A comparison of alternative specifications, *Journal of Business and Economic Statistics*, 21, 368-382.

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

Talukdar, D., K. Sudhir and A. Ainslie (2002), Investigating new product diffusion across products and countries, *Marketing Science*, 21, 97–114.

Tellis, G.J. (1988), Advertising exposure, loyalty and brand purchase: A two stage model of choice, *Journal of Marketing Research*, 25, 134-144.

Tellis, G.J., R. Chandy and P. Thaivanich (2000), Which ad works, when, where, and how often? Modeling the effects of direct television advertising, *Journal of Marketing Research*, 37, 32-46.

Tellis, G.J. and P.H. Franses (2004), The optimal data interval for econometric models of advertising, Review requested by *Marketing Science*.

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

van Nierop, E., D. Fok and P.H. Franses (2002), Sales models for many items using attribute data, ERIM Report Series Research in Management ERS-2002-65-MKT, Erasmus University Rotterdam

Van Oest, R.D., R. Paap and P.H. Franses (2002), A joint framework for category purchase and consumption behavior, Tinbergen Institute report series TI 2002-124/4, Erasmus University Rotterdam

Wedel, M. and W.A. Kamakura (1999), *Market segmentation : conceptual and methodological foundations*, Boston: Kluwer Academic Publishers.

Wieringa, J.E. and C. Horvath (2005), Impulse response analysis of structural vector autoregressive log-transformed time series, *International Journal of Forecasting*, to appear.

Wittink, D.R., M.J. Addona, W.J. Hawkes and J.C. Porter (1988), SCAN*PRO: the estimation, validation, and use of promotional effects based on scanner data, Working paper, AC Nielsen, Schaumburg, Illinois.