State Space Decomposition of Scanner Sales into Trend, Day-of-the-Week Effect, and Multiple Exogeneous Effefcts

Fumiyo N. Kondo

May 13, 1998

Doctor of Philosophy

Department of Statistical Science

School of Mathmatical and Physical Science

The Graduate University for Advanced Studies

Contents

Abstract

Promotion affects short-term retail sales, of which 5 major mechanisms are summarized by Blattberg and Neslin (1990) as brand switching, purchase acceleration, store switching, category expansion, and repeat purchasing. The authors develop a new methodology for decomposing sales into long-term component of baseline sales, cyclical component of day-of-the-week effect, and short-term components of brand substitution and category expansion in order to investigate simultaneously the impacts of retail price promotions on consumer sales. The short-term components are consisted of brand substitution and category expansion including store switching. The methodology is based on a state space model, which is in a line of Bayesian vector state space model and has abilities to specify not only a structure of time series such as ARMA models, but also a cross-sectional structure, further incorporating input-output relationship between prices and sales. The application of the methodology to the daily store level scanner sales data of milk has shown its effectiveness.

Chapter 1

1 Introduction

1.1 Motivation and Background

Store level scanner data contain a large amount of micro level data of sales and prices, which provide good opportunities for a researcher to do a research from various points of view. On the other hand, huge amount of data demands a lot of data handling time. Moreover, there exist various levels of un-interested noise among signals, so that there is always a danger in which a researcher loses his or her way for a analysis. This kind of problems became already known facts among researchers involved in scanner data analysis and are summarized in McCann and Gallagher (1990).

Similarly, D. M. Hanssens et al. (1990) raised the folowing questions in their book, addressing issues of dynamic model specification and data interval on the analysis of time series and cross-sectional data:

1) How do we ensure that the model distinguishes between lagged marketing (effects) and lagged sales effects?
2) What is the best data interval to use in measuring marketing effects? If the grid is too coarse (e.g., annual data), we may not pick up the dynamics of the marketing system. If the grid is too fine, we may be lost in irrelevant data fluctuations.
3) Are the results of a macromodel consistent with those of a micromodel? If not, which one should the modeler believe?
4) Suppose we only have macrodata but wish to examine a microresponse model; Can we infer microbehavior from macrodata?

Conventional models for market responses are largely devided into econometric (regression) model and time series model. A more general approach to combine econometric and time series techniques, ETS in their word, are getting increasing attention. Hanssens and others suggest an ETS approarch rather than a traditional econometric approach to solve problems in dynamic model specification and data interval.

Micro level data are elements of the aggregated macro level data and form a certain structure as a group. Micro level data reflect macro behavior of the aggregated total as well as micro behavior of individual elements. Lets think about what derives micro and macro behavior of scanner data by considering the types of models which are frequently used. For example, when we look at the sales data of products handled at a large supermarket, the following 4 levels of aggregation show different movements against each other:

1) the sales of each SKU (stock keeping unit[1])

2) the sales of each brand

3) the sales of a category

4) the sales of a store total

In practice, 2 different types of models are usually fitted depending on the level of aggregation. In order to estimate the effects of promotion and to forecast weekly sales of a product by using the 1st or 2nd level of aggregation, econometric models such as (log) linear regression models with prices and/or other explanatory variables of promotion are frequently fitted. Regading the 4th level of aggregation, a monthly sales of store total, a time series model with trends and seasonality is often used. The form of the econometric model indicates that present micro level data have a correlation mainly with other present explanatory variables and that of the time series model indicates that the present aggregated data have a correlation mainly with the past aggregated data. Hense, we can assume the behavior of micro level data is mainly derived by the impacts of explanatory variables and those of the aggregated data comes from the past habitual patterns. However, if you think about data generating process, the data in the 4th

[1] the smallest unit stocked on the shelf of a retail store

level of aggregation is generated just by aggregating the data in the former case in the direction of time and products. Therefore, it is expected to exist a mechanism such that an aggregation process of micro level data suppresses large variances existed in the original data, leading to smaller variances in the aggregated data.

Scanner data are always available as multivariate data and there appeared to exist a relatively reasonable basis to assume a multi-level structure such as between SKUs within a brand, between brands within a category, between categories in a store total, or cyclical structure such as days in a week, weeks in a year. our approach is to try to organize micro level data via a unified model with multi-level and cyclical structure as much as possible, by using individual information and group information at the same time for the forecasting. This approach guarantees the results of macro level model are consistent with those of micro level model, so that results are always comparable among elements and the total. If computation capacity permits, an increase in the level of hierachy makes possible to estimate a higher multi-level structure, for example, by specifying the 1st stage elements, the 2nd stage elements that are the 1st stage subtotals, and the grand total. This is a great advantage because accumulated empirical results are always useful to understand the movements of elements and the entire structure. Our research is positioned as a unified method of ETS.

1.2 Literature Review in Combined Methodology of Econometric and Time Series Model in Marketing Research

There are not many applications which take an approach to systhesize econometric and time series model, i.e., an approach that has both time-varying aspects and an input-output relationship, but roughly four methodologies have been used in the analysis of market response. Those are distributed lag models, intervention analysis, transfer function analysis, and a combination of OLS and ARMA model. The following is the application examples of the first 3 models:

Distributed lag models

Bass and Clarke (1972) used distributed lag models to extend a popular Koyck model, in which only monotonic (geometrical) decays of the impact of advertising on sales are allowed and measured the length of the effect of advertising.

Transfer function analysis

Helmer and Johansson (1977) studied sales response to advertising using the Lydia Pinkham data and found that transfer function models had greater forecast accuracy than certain conventional regression models.

Adams and Moriarty (1981) compared the best available regression model to a transfer-function model on an advertising-sales relationship of a product, and concluded that transfer-function analysis are superior to the best regression model.

As well as lagged effects, Doyle and Saunders (1985) studied lead effects, i.e., the prepromotion dip in sales examined department store sales as a function of six marketing instruments:leaflets, display, press, TV advertising, price promotion, and commision structure for sales presonnel.

Intervention analysis

Box and Tiao (1975) studied the intervention effects on a response variable in the presence of dependent noise structure, giving two examples: (1) step intervention effect of the traffic diversion by the opening of the Golden State Freeway and that of the enforcement of a new law on the reduction of hydrocrabons in Los Angeles; (2) step intervention effect of Phase I and Phase II controls on monthly rate of inflation of consumer price index.

Wichern and Jones (1977) investigated the effect of a firm's promotion of the American Dental Association endorsement on market shares of Crest toothpaste by using intervention analysis.

Leone (1987) proposed an intervention analysis procedure to forecast the impact of changes (interventions), showing 2 examples of the impact of a change in advertising on a firm's market share and a price deal effect on sales performance.

Our model is different from the four methodologies, particularly in terms of using original data instead of detrended data, thus accomodating a capacity to directly deal with non-stationary data and further we decompose sales simultaneously into trend, seasonality, and multiple effects caused by exogeneous variabels.

1.3 Price Dynamics Model with Substitution and Cyclical Structure

Blattberg and Neslin (1990) defined the 4 mechanisms of brand switching, repeat purchasing, purchase acceleration, and category expansion as follows:

1) Brand switching

   The consumer is induced to purchase a different brand from that which would have been purchased had the promotion not been available.

2) Reapeat purchasing

   The consumer's probability of buying the brand again in the future is influenced by purchasing the brand on promotion.

3) Purchase acceleration

   The consumer's purchase timing or purchase quantity is changed by the promotion.

4) Category Expansion

   The consumer's total consumption of the product category is increased by the promotion.

With respect to store switching, they explained as follows: "A consumer may respond to a retailer promotion by switching stores, that is, shopping at a different store than the store he or she would have had the retailer promotion not been offered."

There have been many articles publised on the subject of substitution (switching) on sales promotion. Some of them are as follows:

Blattberg and Wisniewski (1989) studied 4 categories of flour, margarine, bathroom tissue, and cannned tuna on brand switching, caused by price promotions. They claimed their results showed asymmetric pattern of price competition: higher-priced, higher quality brands steal share from other brands in the same price-quality tier, as well as from brands in their below, but, lower-priced, lower quality brands can not steal share from higher-priced, higher quality brands.

By using consumer scan panel data[2], Sunil (1988) reports that for the examined category of coffee, more than 84% of substitution is due to brand substitution, less than 14% is due to purchase accerelation and less than 2% is due to stockpile effect.

Kumar and Leone (1988) studied store-level scanner data of the disposable diaper category on the effects of sales promotion and found that although there is significant promotion-induced store substitution between neighboring stores, within-store substitution rates were two to three times greater.

[2]Consumer panel data provide information on individual household purchses, whereas store level data contain all sales in a given store or collection of stores over a period of time. The scanner data are collected by a scanning equipment placed at each checkout counter of a retail store and is connected to a computor. As each item is scanned, the count is updated for the particular code such as JAN code or UPC code on the purchased package. Store data contain aggregate sales from all consumers shopping in the store(s). Consumer panel data contain household-level sales of a panel of consumers who present a special card at the check-out counter and are a subset of all consumers shopping in the store.

Walters (1991) examined on store-level scanner data both within-store and between-store substitution and complementary effects. Store substitution rates were very low compared to brand switching within a store.

The substitution effects are the most easily observed phenomena in micro level scanner data, so that we assume subsitution is a key mechanism which derives large variances in the micro level original data and use it to organize scanner dala in a structural model. There are several causes to induce subsitution, but temporary price reduction can naturally be thought of the largest factor to heighten a purchase desire by a consumer. Meanwhile, a significant reduction on the price of a brand may not lead to an increase in the purchases of the category total. This "zero-sum" effect in the direction of brand occurs when the lower priced brand is purchased as a substitute of the competitive brand, but not as an incremental demand of the category.

If only a brand substitution occurs, the explanatory variable of price has an enormous influence on the brand sales data, but has no influence on

the aggregated data of the category. This leads to an expectation that such a phenomenon can exist that, for brand level, sales data have a strong correlation with price, but not for category level. An increase in sales also arises due to the fact that consumers substitute from future sales as consumers advance their purchases in time, for example, by purchasing a promoted brand in the promotion period twice as much as in non-promotion period. Furthermore, an increase in sales may come from a category expansion, in which a substitution from other stores can be confounded.

Research on the effectiveness of a sales promotion by decomposing the sales "bump" during the promotion period into incremental sales due to brand switching, purchase time acceleration, and stockpiling was done on consumer scan panel data by Sunil (1988). Also, there is a major research on brand switching effect of promotion on store level scanner data by Blattberg and Wisniewski (1989) by the method of OLS.

Our research is to develop a new methodology to decompose store level sales into long-term component of baseline sales (trend), cyclical day-of-the-week effect components, and short-term incremental sales components of brand substitution and/or category expansion caused by price promotion, simultaneously, utilizing a Bayesian vector state space model with time-varying parameters. Unlike the study done by Sunil (1988) on consumer scan panel data, we can do direct forcasts of the brand sales and the category total for a store since we utilizes store level scanner sales data. Also, this method can provide up-to-date forecasts by using time-varying parameters. The most significant difference from conventional models is that this model can avoid "pooling" of original data in order to obtain parameter estimates. The model utilizes the method of type II maximum likelihood having a Bayesian procedure, which can deal with a model with more parameters than the number of data points. These are methodological advantages against a constant parameter study on pooling data done by Blattberg and Wisniewski (1989).

In order to unify the terminology regarding "zero-sum" effect, "brand switching" and "purchase acceleration and stockpile" effect are hereafter refer to as "brand substitution" and "temporal substitution [3] ", respectively.

[3] There may be a different opinion on calling "temporal substitution". For example, Frand and Massy (1971) described the term, "substitution", for reduction in both current and future sales for competitior products, and the term, "displacement", for reduction of subsequent nonpromotional period retail sales.

Chapter 2

2 Sales Decomposition Model

2.1 Basic Model

The sales of brands are represented by k-variate time series $y(n) = (y_1(n), \dots, y_k(n))^T$, $(n = 1, \dots, N)$ and are assumed to be able to decompose as follows:

$$y(n) = t(n) + d(n) + x(n) + w(n), \tag{2.1}$$

where $t(n)$, $d(n)$, $x(n)$ and $w(n)$ represent long-term baseline sales component, cyclical day-of-the-week effect component, short-term component of substitution/category expansion, and observation noise, respectively.

The observation noise $w(n)$ obeys the following normal distribution:

$$w(n) \sim N(\mathbf{0}\ , \Sigma_w), \quad \Sigma_w = \begin{pmatrix} \sigma^2_{w_{11}} & \cdots & \sigma^2_{w_{1k}} \\ \vdots & \ddots & \vdots \\ \sigma^2_{w_{k1}} & \cdots & \sigma^2_{w_{kk}} \end{pmatrix}. \tag{2.2}$$

Short term component is decomposed into substitution and category expansion. Substitution component may be further decomposed into "brand substitution", "temporal substitution", and "store substitution". Although past researches such as Kumar et al. (1988) suggest the existance of "store substitution" effect, the substitution components in this paper are limited to "brand substitution". Due to the difficultity in data availability for competitive stores, "Store substitution" is instead included in category expansion. For "temporal substitution", only the model framework is given in the APPENDIX, for future data analysis.

Therefore, the short term component is decomposed as follows:

$$x(n) = g(n) + s(n) + z(n), \tag{2.3}$$

where $g(n)$ and $s(n)$ represent "brand substitution", "temporal substitution", and "category expansion", respectively.

Decomposing price promotion effect into each component has a benefit in understanding which direction among brand, category total (including store), and time, a price promotion has effectively affected. Except time effect, the concept of this decomposition is virtually the same as that of empirical marketing effect decomposition of "competitive effect", "primary sales effect", and "primary demand effect". Competitive effect is the effect to increase its own sales and decrease those of competitors. Primary sales effect is the effect to increase its own sales without affecting those of competitors. Primary demand effect is the effect to increase both the own sales and the competitors' sales. A taxonomy is given by Shultz and Wittink (1976).

For each brand i, the sales can be written as

$$y_i(n) = t_i(n) + d_i(n) + g_i(n) + s_i(n) + z_i(n) + w_i(n), \quad i = 1, \ldots, k. \tag{2.4}$$

By summing the sales of each brand, the category total sales can be expressed as

$$\begin{aligned} y_\bullet(n) &= t_\bullet(n) + d_\bullet(n) + g_\bullet(n) + s_\bullet(n) + z_\bullet(n) + w_\bullet(n) \\ &= t_\bullet(n) + d_\bullet(n) + s_\bullet(n) + z_\bullet(n) + w_\bullet(n), \end{aligned} \tag{2.5}$$

where $y_\bullet(n) = \sum_{i=1}^{k} y_i(\mathrm{n})$ and $g_\bullet(n) = \sum_{i=1}^{k} g_i(n) = 0$. Thus, for the category total, the term for brand substitution is vanished. This can be the main reason that an aggregation in terms of brand removes a large amount of variances which existed in each brand data under the competitive structure of "brand substitution". By the same token, if there exists a "temporal substitution" or a "store substitution", an aggregation in terms of adjacent period or store will remove variances existing in the original data by the portion of substitution effect in time or store.

Figure 1 is the graphs of the data analyzed in this research. The first to the fourth graphs in the figure show movements of major 4 brands in the period of 2 years. The last graph is for the sum of the 4 brands and shows a brand substitution mechanism in which an aggregation of 4 brands removed variances that existed in the orginal daily scanner data and exhibited an obvious seasonality.

2.2 Long-term Local Polynomial Baseline Component

Baseline sales for a brand are defined as the sales without any sales promotions and baseline component in our model is characterized as habitual repeat purchasing that is not affected by sales promotion. We assume that long-term effects of repeat purchasing or those of category expansion by sales promotion does not exist or negligible. Although no empirical researches reporting the existence of repeat purchasing effects or category expansion effects in baseline sales caused by sales promotion are known to us, if the effects of those two components actually exist, they are confounded to the baseline sales of brands. There might exist explanatory variables which affect long-term component such as TV-advertisement, but it is not considered in this paper.

We deal with an old product category that is already existing in the market place. A very recently entered new product category into the market, which is in the field of a diffusion model, is not dealt in this paper. There is a recent study on new product diffusion model with a Kalman filter (Xie et al. (1997)).

Local polynomial baseline component, $t(n) = (t_1(n), \ldots, t_k(n))^T$ is represented by the following l th order stochastic difference equation:

$$\Delta^l t(n) = v_l(n), \tag{2.6}$$

where the system noise $v_{l_i}(n) = (v_1(n), \ldots, v_k(n))^T$ obeys the following Gaussian white noise

$$v_l(n) \sim N(\mathbf{0}\ , \Sigma_l), \quad \Sigma_l = \begin{pmatrix} \sigma^2_{l_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma^2_{l_k} \end{pmatrix}, \tag{2.7}$$

and Δ denotes the difference operator defined by $\Delta t(n) = t(n) - t(n-1)$. Here, the smoothness of the trend are controlled by the variance of the system noise, σ^2_l, and the order of l (Kitagawa and Gersch (1984)).

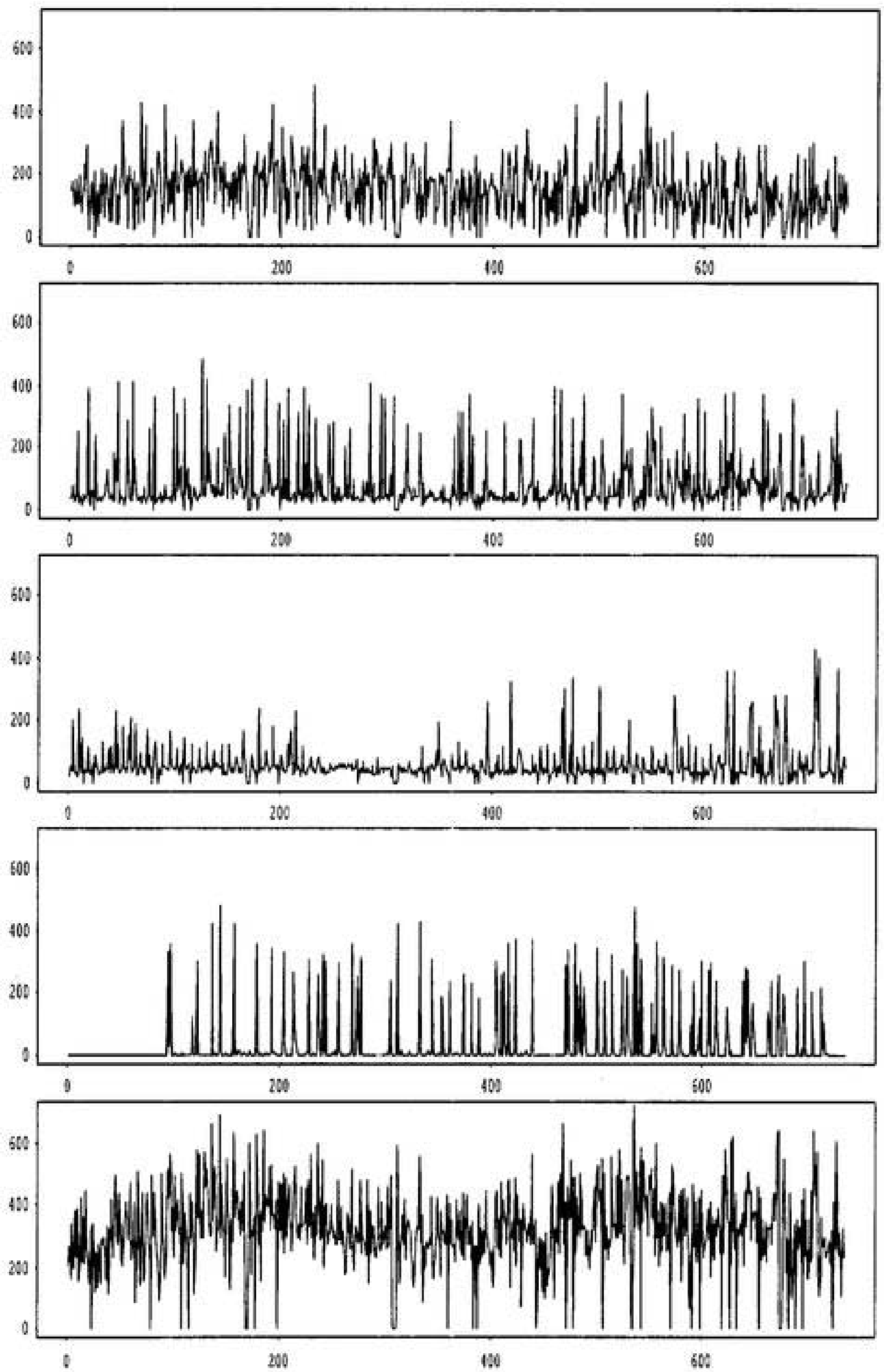


Figure 2.1: 4 Brands and the Sum of Daily Scanner Milk Data
Variance for 4 Brands and the Sum (7584; 7930; 3138; 8059; 17213)

2.3 Cyclical Day-of-the-Week Component

Cyclical day-of-the-week component is a component which repeats similar patterns on the same day within a week. In one week cycle, there exist 7 observations and cyclical component may be expressed by

$$d(n) \approx d(n-7), \tag{2.8}$$

where $d(n) = (d_1(n), \ldots, d_k(n))^T$, is a k-variate day-of-the-week component.

When cyclical component is used together with trend component, the following representation for cyclical component is used instead of the representation in equation (2.8) in order to guarantee uniqueness of the decomposiotion (refer to Kitagawa and Gersch (1984)):

$$\sum_{j=0}^{6} d(n-j) = v_d(n), \tag{2.9}$$

where the system noise $v_d(n) = (v_1(n), \ldots, v_k(n))^T$ obeys the following Gaussian white noise and corresponds to the change of the day-of-the-week pattern:

$$v_d(n) \sim N(\mathbf{0}\ , \Sigma_d), \quad \Sigma_d = \begin{pmatrix} \sigma^2_{d_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma^2_{d_k} \end{pmatrix}. \tag{2.10}$$

Note that for small $v_d(n)$, $d(n)$ behaves as periodic function (Kondo and Kitagawa (1998)).

2.4 Short-term Two Explanatory Variable Effect Components for Incremental Sales

There are several vehicles for sales promotions such as price promotion, display, advertisements, coupons, and so on. Price promotion is considered to be the largest factor to induce consumers a desire for incremental purchases. Therefore, we selected price promotion as the representative explanatory variable. If data are available for other sales promotions, they can be included in the model.

Exogeneous effect components are consisted of components of brand substitution and category expansion. Temporal substitution remains future research. For each component, how sales are related to prices must be determined and a regression model plays an important role on this subject. Each component is a response variable to explanatory variables of "price function" which are defined in the next section.

2.4.1 Price function

Blattberg and Neslin (1990, p.235) view sales promotion, including price promotion, as short-term pulses "intervening" with the normal progression of the sales of time series. We share the view by considering price promotions are effective to cause short-term pulses of sales. The relationship between consecutive inputs and the resultant consecutive effects can be dealt with transfer function analysis. Intervention analysis treats the cases that inputs are temporary pulses or permanent step change. However, the both do not treat consecutive inputs and the resultant intermittent effects or temporary pulse effects. The explanatory variable of price is consecutive inputs and the resulting effects on sales are intermittent effects through information processing within consumer's brain. Therefore, in order to treat this problem as a regression analysis, there is a necessity to transform consecutive inputs into intermittent inputs. The use of price cut level as inputs instead of price itself is the first step to produce the intermittent inputs. Further, a reduction of price does not always yield pulse-like incremental sales. This may be due to too weak inputs to produce any effects or a difference in competitive strength among brands or empirically suggested "deal decay". Therefore, as the second step to synchronize inputs to the occurrences of intermittent incremental sales, we considered an interfering mechanism in the occurrences of price promotional effects by competitors or a wearing-out process on promotional effects as time passes. In summary, the relationship between prices and sales, i.e., whether intermittent incremental sales are produced by a price reduction or not must be considered and a model describing an input-output relationship plays an important role on this subject. The condition that pulse-like incremental sales are produced by a price reduction can be defined and referred to as a "price function" hereafter.

Let $u(n) = (u_1(n), \ldots, u_k(n))^T$ be a k-variate vector function of price such as "price cut" as follows:

$$u(n) = f(p(n)) \leq 0, \tag{2.11}$$

where $p(n) = (p_1(n), \ldots, p_k(n))^T$ is a k-variate vector of actual prices of brands within a product category. Price function is determined from a set of univariate models of brands. The form of price function considered is a decreasing function from the origin of zero, at which the sales level corresponds to that of the trend. The larger the absolute value of price function, the greater the potential of incremental sales becomes.

The level of data aggregation can change the appearance of the relationship between prices and sales, so that the price function of daily aggregated data can be quite different from weekly aggregated data if prices vary largely among days within a week. Considering the practical situation that a consumer selects a purchasing brand by comparing the actual price and the regular price of a brand, and by comparing prices among shelf-stocked competitive brands, the first simple choice of price function is in a form described with differences in prices. Therefore, we considered 5 candidates of instantaneous price function at first. For the first 4 price functions, used is the difference of actual price from the maximum price during the entire periods, $\tilde{u}(n) = (\tilde{u}_1(n), \ldots, \tilde{u}_k(n))^T$, which is defined as follows:

$$\tilde{u}_i(n) = p_i(n) - \max_{n \in \{1,\ldots,N\}} p_i(n) \leq 0, \quad i = 1, \ldots, k. \tag{2.12}$$

Since we do not have the information on regular prices, we use the maximum price as a substitute for a regular price. For the 5th instantaneous price function, a direct comparison among price levels is considered. The followings are the 5 price functions:

f_1) a price cut of a brand from the maximum price during the entire periods

$$u_i(n) \;=\; \tilde{u}_i(n). \tag{2.13}$$

f_2) a relative price cut to those of competitors

(incremental effects are yielded only when the price cut is greater than that of its competitors)

$$\begin{aligned} u_i(n) &= \tilde{u}_i(n) \qquad if \;\; \tilde{u}_i(n) \leq \min_{j\in\{1,\ldots,k\}, j\neq i} \tilde{u}_j(n) \\ &= 0 \qquad\quad\;\; \text{otherwise.} \end{aligned} \tag{2.14}$$

f_3) the maximum price cut among brands

(incremental effects are yielded under the same condition as f_2, but the input level is determined by the difference of the price cut from that of its competitors)

$$\begin{aligned} u_i(n) &= \tilde{u}_i(n) - \min_{j\in\{1,\ldots,k\}, j\neq i} \tilde{u}_j(n) \qquad if \;\; \tilde{u}_i(n) \leq \min_{j\in\{1,\ldots,k\}, j\neq i} \tilde{u}_j(n) \\ &= 0 \qquad\qquad\qquad\qquad\qquad\quad \text{otherwise.} \end{aligned} \tag{2.15}$$

f_4) a price cut with a lower and an upper threshold

(incremental effects are yielded only when the price cut is within a certain range)

$$\begin{aligned} u_i(n) &= \tilde{u}_i(n) \qquad if \;\; Lth \leq -\tilde{u}_i(n) \leq Uth \\ &= 0 \qquad\quad\;\; \text{otherwise,} \end{aligned} \tag{2.16}$$

where $0 < Lth < Uth$ determines lower and upper threshold.

f_5) a relative price to those of competitors

(incremental effects are yielded only when the price is the lowest among competitors)

$$\begin{aligned} u_i(n) &= -1 \qquad if \;\; p_i(n) \leq \min_{j\in\{1,\ldots,k\}, j\neq i} p_j(n) \\ &= 0 \qquad\;\;\; \text{otherwise.} \end{aligned} \tag{2.17}$$

Further, by using the best function selected among f_1 through f_5, the following two non-linear functions, a deal decay function and an exponential function, were included due to empirical consideration and the establishment of input level saturation point. In the below, $f_b[\,\cdot\,]$ denotes the best selected function among f_1 – f_5.

f_6) deal decay function

(promotion effect decays in time if promotion runs more than one time period)
Deal-decay function is considered to incorporate empirical result that promotional effects decay if promotion continues more than one time period as a result of consumers' response pattern. Further, because of physical obstruction in sales such as out-of-stock, we incorporated a mechanism to reset to the initial effect level.

$$u_i(n) = \exp\Big\{-\gamma(n - n_0)\Big\} f_b[u_i(n)], \tag{2.18}$$

where $\gamma \geq 0$ is a constant parameter and n_0 is the first period of promotion run and is reset to n when $n - n_0$ becomes greater than η (a positive integer) under the hypothesis that deal decay stops somewhere within a short span.

f_7) exponential function

(inputs have a saturation point defined by an exponential function)

$$u_i(n) = \exp\Big\{\nu f_b[u_i(n)]\Big\} - 1, \tag{2.19}$$

where $\nu \geq 0$ is a constant parameter.

2.4.2 Brand Substitution Component Model

Brand substitution component model describes a component which has large variaces for each brand, but vanishes if they are summed by brand. Therefore, it is characterised as "zero-sum effect" in the direction of brand.

Let us assume a brand substitution component, $g(n) = (g_1(n), \ldots, g_k(n))^T$, is represented by

$$g(n) = B(n)u^g(n), \tag{2.20}$$

where $u^g(n)$ is defined in the below as a price function for brand substitution model and $B(n)$ represents time-varying coefficients of $u^g(n)$, as follows:

$$B(n) = \begin{pmatrix} b_{11}(n) & \cdots & b_{1k}(n) \\ \vdots & \ddots & \vdots \\ b_{k1}(n) & \cdots & b_{kk}(n) \end{pmatrix}. \tag{2.21}$$

The coefficients, $b_{ij}(n)$ are assumed to be flexible time-varying parametes in order to respond to changes in the market place and also absorb seasonality. Since coefficients are assumed to have mild changes, they can be expressed as a locally constant component with the following 1st order stochastic difference equation:

$$b_{ij}(n) - b_{ij}(n-1) = v_{b_{ij}}(n) \quad i, j = 1, \ldots, k, \tag{2.22}$$

with an environmental system noise $v_{b_{ij}}(n)$, obeying the following normal distribution,

$$v_{b_{ij}}(n) \sim N(0, \sigma^2_{b_{ij}}), \quad \sigma^2_{b_{ij}} = \sigma^2_b(\sigma^2_{t_i} / \sum_{j=1}^{k} \sigma^2_{t_j}), \tag{2.23}$$

where, $\sigma^2_{b_{ij}}$ is a product of σ^2_b, a common parameter and a weight determined by $\sigma^2_{t_i}$, the variance of system noise for baseline component for a brand i.

The competitive structure of brand substitution can be expressed as a constraint of the equation (7.81) and is represented by

$$b_{ii}(n) \leq 0, \quad b_{ij}(n) \geq 0, (i \neq j), \quad \sum_{i=1}^{k} b_{ij}(n) = 0. \tag{2.24}$$

These constraints have a role to define a 100% basis of the sales for brand substitution, confining price promotion effect into the one for brand substitution.

2.4.3 Category Expansion Component Model

Category expansion component is the remaining incremental sales effects after removing the ones for substitution, which contributes to a net increase in the sales of the category.

Let us assume a category expansion component, $z(n) = (z_1(n), \ldots, z_k(n))^T$, $z_i(n) \geq 0, \quad i = 1, \ldots, k$ is represented by

$$z(n) = \Lambda(n)u^g(n), \tag{2.25}$$

where $u^g(n)$ is the same price function as the one for brand substitution and $\Lambda(n)$represents time-varying coefficients of $u^g(n)$ as follows:

$$\Lambda(n) = \begin{pmatrix} \lambda_{11}(n) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_{kk}(n) \end{pmatrix}. \tag{2.26}$$

Category expansion is a non-competitive effect, i.e., a primary sales effect and/or a primary demand effect, so that own sales and competitors sales must increase or be zero. Therefore, only diagonal elements of coefficients must be non-zero, because if off-diagonal elements that are effects from competitors prices were non-zero, the sales of some brands can decrease.

The coefficient, $\lambda_{ii}(n)$, is assumed to be a flexible time-varying parameter as in brand substitution component, being expressed as a locally constant component with the following 1st order stochastic difference equation:

$$\lambda_{ii}(n) - \lambda_{ii}(n-1) = v_{\lambda_{ii}}(n) \quad i = 1, \ldots, k, \tag{2.27}$$

where a system noise, $v_{\lambda_{ii}}(n)$, obeys the following normal distribution,

$$v_{\lambda_{ii}}(n) \sim N(0, \sigma_{ii}^2), \quad \sigma_{ii}^2 = \sigma_b^2(\sigma_{t_i}^2 / \sum_{j=1}^{k} \sigma_{t_j}^2), \tag{2.28}$$

where, σ_{ii}^2 is a product of σ_b^2, a common parameter and a weight determined by $\sigma_{t_i}^2$, the variance of system noise for baseline sales for a brand i.

The conditions of category expansion can be expressed with coefficients λ_{ii} as follows:

$$\lambda_{ii}(n) \leq 0. \tag{2.29}$$

Chapter 3

3 Estimation and Identification of Model

Liu and Hanssens (1981) used a Bayesian approach with Kalman filter to analyze the relationship between sales and price on inexpensive gift brands. They analyzed cross-sectional time series data with a time-varying regression model. The type of variation[4] is sequential stochastic parameter variation. As the domain of variation, their research considered only time variation, and no cross-sectional variation with the same coefficient across cross-section.

Our analysis also follows Bayesian vector state space model on cross-sectional and time series data with sequential stochastic parameter variation, but facilitates cross-sectional variation as well as time variation. We focuss on sales decomposition on substitution and category expansion effects caused by retail price promotions, so that our model deals with decomposition into trend, cyclical component, and multiple effects caused by explanatory variables.

Taking a Bayesian approach is due to the reason that the used daily scanner data contains complicated responses to complex market environments such as new entries of brands, changes in marketing strategies, consumer tastes, competition, and so on. A Bayesian approach provides flexible means to the analysis of such complex data, satisfying the requirements for the estimation of many parameters, time-varying coefficients, and the specification of structures.

[4] A. R. Wildt and R. S. Winer (1983) summarized time-varying parameter models in marketing applications. They classified variable-parameter models on the basis of the domain of variation, the type of variation, and the extent of variation. The type of variation is classified due to systematic variation caused by observable variables or stochastic variation. Within stochastic variation, there are random variation and sequential variation. The distinction between them is that sequantial models hypothesize specific stochastic processes that coefficients follow over time, whereas random models hypothesizd deviations from a mean vector resulting from strictly random variations.

3.1 Smoothness Priors Determined by Maximum Likelihood in Bayesian Model

In an analysis of input-output relationship on econometric time series, Shiller(1973) introduced the notion of "smoothness priors", having "smoothness" constraints on distributed lags in a difference equation: A single trade-off parameter determines the trade-off between infidelity of the model to the data and infidelity of the model to the smoothness constraints. A similar concept appeared in Whittaker(1923), addressing a problem of the estimation of a smooth trend. The trade-off parameter was determined subjectively untill Akaike (1980a,1980b) formulated an objective method in a quasi-Bayesian approach.

Bayes' law provides a mathematical procedure for updating prior information on a parameter in a distribution family, producing a posterior distribution of the parameter. In Bayesian inference, Lindley and Smith (1972) claimed that prior and posterior probabilities can be interpreted as subjective probabilities, calling priors as hyperparameters. Marits and Lwin (1989) introduced empirical Bayes method, in which previous data are used to get an estimate of the prior distribution. Harrison and Stevens (1976) used a dynamic linear model, a regression model and a multi-process model to obtain posterior updating the prior of an estimated parameter.

Akaike proposed the method of choosing the trade-off parameter, or hyperparameter in a Bayesian terminology, by maximizing the likelihood of a Bayes model, which are defined by mixing the data distribution with prior weights to yield a marginal likelihood computation among several candidates of prior distributions. The caluculation of the marginal likelihood requires intensive computation, of which burdan Kitagawa and Gersch (1983) eased by employing a state space representation of the model, using the recursive algorithm of Kalman filtering. Kalman filter algorithm theoretically provides a method for the exact maximum likelihood estimation of the model. Also, prediction on the dependent variable is effectively solved as described in the following sections.

3.2 State Space Model Representation

The time series model explained so far is given by

$$y(n) = t(n) + d(n) + x(n) + w(n). \tag{3.30}$$

This model can be also expressed in a form of a general state space model:

$$y(n) = H(n)\alpha(n) + w(n) \quad \text{(observation model)} \tag{3.31}$$

$$\alpha(n) = F\alpha(n-1) + Gv(n) \quad \text{(system model)} \tag{3.32}$$

or

$$\begin{pmatrix} t(n) \\ d(n) \\ b(n) \\ \lambda(n) \end{pmatrix} = \begin{pmatrix} I_k & & & \mathbf{0} \\ & F_D & & \\ & & F_B & \\ \mathbf{0} & & & F_\lambda \end{pmatrix} \begin{pmatrix} t(n-1) \\ d(n-1) \\ b(n-1) \\ \lambda(n-1) \end{pmatrix} + \begin{pmatrix} v_t(n) \\ v_d(n) \\ v_b(n) \\ v_\lambda(n) \end{pmatrix}, \tag{3.33}$$

$$\begin{pmatrix} w(n) \\ v(n) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_w & 0 \\ 0 & Q \end{pmatrix} \right), \tag{3.34}$$

where $\alpha(n) = (t(n), d(n), b(n), \lambda(n))^T$, $w(n)$, and $v(n)$ are state vector, observation noise and system noise, respectively. Each component of the state vector $\alpha(n)$, system noise $Gv(n)$, matrices F and $H(n)$ are specified in the section 3.6.

3.3 State Estimation by Kalman Filter

The state space representation of our model places the problem to estimate the state vector. Kalman filter estimates the state by evaluating the distribution of $\alpha(n)$ given the observation $Y_l \equiv (y_1, \ldots, y_l), l = 1, \ldots, N$ and the initial values $\alpha_{0|0}$ and $V_{0|0}$.

The following is the recursive process of Kalman filter.

1) Prediction of State Vector

   By using the equation (3.35), (3.36) of the well known Kalman filter formula in the below, predict the next state $\alpha_{1|0}$ and $W_{1|0}$ with initial condition $\alpha_{0|0}$, $W_{0|0}$.

2) Filtering of State Vector

   Update $\alpha_{1|1}$, $W_{1|1}$ by calculating the equation (3.38), (3.39) with new observation y_1.

   Continue the process 1) and 2) until the whole calculation of $\alpha_{l|l-1}, \alpha_{l|l}, W_{l|l-1}, W_{l|l}, l = 1, \ldots, N$ has completed.

3) Smoothing of State Vector

   Obtain the state vector $\alpha_{N-1|N}$, $W_{N-1|N}$ by using the fixed interval smoother of the euqation (3.41), (3.42). Continue the process 3) until the calculation of $\alpha_{N-1|N}, \ldots, \alpha_{1|N}$, $W_{N-1|N}, \ldots, W_{1|N}$ has completed.

   Kalman Filter formula are given as follows:

1) Prediction (Time-Update) Formula

$$\alpha_{n|n-1} = F\alpha_{n-1|n-1} \tag{3.35}$$

$$W_{n|n-1} = FW_{n-1|n-1}F^T + G_nQG_n^T. \tag{3.36}$$

   where the initial conditions $\alpha_{0|0}$ and $W_{0|0}$ are assumed to be given.

2) Filter (Observation-Update) Formula

$$K_n = W_{n|n-1}H_n^T(H_nW_{n|n-1}H_n^T + R_n)^{-1} \tag{3.37}$$

$$\alpha_{n|n} = \alpha_{n|n-1} + K_n(y_n - H_n\alpha_{n|n-1}) \tag{3.38}$$

$$W_{n|n} = (I - K_nH_n)W_{n|n-1}. \tag{3.39}$$

3) Smoothing Formula

$$\begin{aligned}
A_n &= W_{n|n} F^T W_{n+1|n}^{-1} && (3.40)\\
\alpha_{n|N} &= \alpha_{n|n} + A_n(\alpha_{n+1|N} - \alpha_{n+1|n}) && (3.41)\\
W_{n|N} &= W_{n|n} + A_n(W_{n+1|N} - W_{n+1|n})A_n^T. && (3.42)
\end{aligned}$$

3.4 Identification of the Model

The likelihood of the model can be expressed as the factorized conditional distributions with parameter θ as follows:

$$\begin{aligned}
L(\theta) &= f(y_1, \ldots, y_N|\theta)\\
&= f(y_1, \ldots, y_{N-1}|\theta) f(y_N|y_1, \ldots, y_{N-1}, \theta)\\
&= \quad \vdots\\
&= \prod_{n=1}^{N} f(y_n|y_1, \ldots, y_{n-1}, \theta)\\
&= \prod_{n=1}^{N} f(y_n|Y_{n-1}, \theta), && (3.43)
\end{aligned}$$

where $Y_{n-1} \equiv \{y_1, \ldots, y_{n-1}\}$.

The individual terms are given by

$$\begin{aligned}
f(y_n|Y_{n-1}) &= (2\pi)^{-k/2}(|V_{n|n-1}|)^{-1/2} && (3.44)\\
&\quad \times \exp(-0.5(y(n) - y_{n|n-1})^T V_{n|n-1}^{-1}(y(n) - y_{n|n-1})), && (3.45)
\end{aligned}$$

where $y_{n|n-1}$ and $V_{n|n-1}$ are the mean and the variance covariance matrix of the observation y_n, respectively, and defined by

$$\begin{aligned}
y_{n|n-1} &= H_n \alpha_{n|n-1} && (3.46)\\
V_{n|n-1} &= H_n W_{n|n-1} H_n^T + \Sigma_w. && (3.47)
\end{aligned}$$

Here $\alpha_{n|n-1}$ and $W_{n|n-1}$ are the mean and the variance covariance matrix of the state vector given the observations y_{n-1} and can be obtained by the Kalman filter shown in section (3.3).

The distribution of $y(n)$ based on the information up to $n-1$ period obeys the following normal distribution:

$$y(n) \sim N(y_{n|n-1}, V_{n|n-1}), \quad V_{n|n-1} = \begin{pmatrix} \sigma^2_{y_{11}} & \cdots & \sigma^2_{y_{k1}} \\ \vdots & \ddots & \vdots \\ \sigma^2_{y_{k1}} & \cdots & \sigma^2_{y_{kk}} \end{pmatrix}. \tag{3.48}$$

The log-likelihood of the model can be written as

$$\log L(y|\theta) = -0.5(Nk \log 2\pi + \sum_{n=1}^{N} \log |V_{n|n-1}| \tag{3.49}$$

$$+ \sum_{n=1}^{N} (y(n) - y_{n|n-1})^T V^{-1}_{n|n-1} (y(n) - y_{n|n-1})). \tag{3.50}$$

3.5 Dynamic Recursive Model Having Major Brands and Category Total

When we analyze marketing sales data, we seldom do analysis on every SKU. On the other hand, we select interested brands and make the other brands total, and the category total. The very information which reseachers need is on brands and the category total, instead of brands and the other brands total, so that we arrange to estimate and forecast on brands and the other total by using information on brands and the category total.

Consider the situation that we have $k-1$ brands and the other total. Lets produce $\tilde{y}(n)$ by multiplying $y(n)$ by Γ defined in (3.52), so that the equation

$$\tilde{y}(n) = \Gamma y(n) = \Gamma H \alpha(n) + \Gamma w(n) \tag{3.51}$$

is for $k-1$ brands and the category total, where Γ is defined by

$$\Gamma = \begin{pmatrix} I_{k-1} & \mathbf{0} \\ \mathbf{1} & 1 \end{pmatrix}, \quad \Gamma^{-1} = \begin{pmatrix} I_{k-1} & \mathbf{0} \\ \mathbf{-1} & 1 \end{pmatrix}, \quad |\Gamma^{-1}| = |\Gamma| = 1. \tag{3.52}$$

The distribution of $\tilde{y}(n)$ based on the information up to $n-1$ period obeys the following normal distribution:

$$\tilde{y}(n) \sim N(\tilde{y}_{n|n-1}, \tilde{V}_{n|n-1}), \quad \tilde{V}_{n|n-1} = \begin{pmatrix} \sigma^2_{\tilde{y}_{11}} & \cdots & \sigma^2_{\tilde{y}_{k1}} \\ \vdots & \ddots & \vdots \\ \sigma^2_{\tilde{y}_{k1}} & \cdots & \sigma^2_{\tilde{y}_{kk}} \end{pmatrix}, \tag{3.53}$$

where $\sigma^2_{\tilde{y}_{ik}} > 0, \quad \sigma^2_{\tilde{y}_{ki}} > 0, \quad i = 1, \ldots, k.$

The distribution of $y(n)$ based on the information of $\tilde{y}(n)$ up to $n-1$ period obeys the following normal distribution:

$$y(n) \sim N(\Gamma^{-1}\tilde{y}_{n|n-1}, \Gamma^{-1}\tilde{V}_{n|n-1}\Gamma^{-T}). \tag{3.54}$$

The log-likelihood of the model is given by

$$\begin{aligned} \log L(y|\theta) &= -0.5(Nk\log 2\pi + \sum_{n=1}^{N} \log|\Gamma^{-1}\tilde{V}_{n|n-1}(\Gamma^{-1})^T| \\ &+ \sum_{n=1}^{N}(y(n) - \Gamma^{-1}\tilde{y}_{n|n-1})^T(\Gamma^{-1}\tilde{V}_{n|n-1}(\Gamma^{-1})^T)^{-1}(y(n) - \Gamma^{-1}\tilde{y}_{n|n-1})) \\ &= -0.5(Nk\log 2\pi + \sum_{n=1}^{N} \log|\tilde{V}_{n|n-1}| \\ &+ \sum_{n=1}^{N}(\Gamma y(n) - \tilde{y}_{n|n-1})^T\tilde{V}^{-1}_{n|n-1}(\Gamma y(n) - \tilde{y}_{n|n-1})) \end{aligned} \tag{3.55}$$

3.6 Specification of System Model and Matrix H(n) for Observation Model

The system model of the general state space model in (3.33) can be specifically rewritten as follows:

$$\begin{pmatrix} t(n) \\ d(n) \\ b(n) \\ \lambda(n) \end{pmatrix} = \begin{pmatrix} I_k & & & \mathbf{0} \\ & F_D & & \\ & & F_B & \\ \mathbf{0} & & & F_\lambda \end{pmatrix} \begin{pmatrix} t(n-1) \\ d(n-1) \\ b(n-1) \\ \lambda(n-1) \end{pmatrix} + \begin{pmatrix} v_t(n) \\ v_d(n) \\ v_b(n) \\ v_{\lambda(n)} \end{pmatrix}. \tag{3.56}$$

If a state vector of $b(n)$ has no constraint and no decomposition into brand substitution and category expansion is involved, F_B are specified as I_{k^2} and a state vector on $\lambda(n)$ is unnecessary. For the brand substitution plus category expansion model, which has a substitution constraint and a category expansion condition, F_B and F_λ will be explained in section 3.6.2 and section 3.6.4. Each component of $t(n), d(n), b(n), \lambda(n), v_t, v_d, v_b,$ $v_\lambda, F_D, and the matrix \mathrm{H(n)} = \begin{pmatrix} I_k & H_d & H_b & H_\lambda \end{pmatrix}$ in observation model (3.31) are defined as follows:

$$t(n) = \begin{pmatrix} t_1(n) \\ \vdots \\ t_k(n) \end{pmatrix}, \quad d(n) = \begin{pmatrix} d_1(n) \\ \vdots \\ d_1(n-5) \\ \vdots \\ d_k(n) \\ \vdots \\ d_k(n-5) \end{pmatrix}, \quad b(n) = \begin{pmatrix} b_{11}(n) \\ \vdots \\ b_{1k}(n) \\ \vdots \\ b_{k1}(n) \\ \vdots \\ b_{kk}(n) \end{pmatrix}, \quad \lambda(n) = \begin{pmatrix} \lambda_{11}(n) \\ \vdots \\ \lambda_{1k}(n) \\ \vdots \\ \lambda_{k1}(n) \\ \vdots \\ \lambda_{kk}(n) \end{pmatrix} \tag{3.57}$$

$$v_t(n) = \begin{pmatrix} v_{t_1}(n) \\ \vdots \\ v_{t_k}(n) \end{pmatrix}, \quad v_d(n) = \begin{pmatrix} v_{d_1}(n) \\ 0 \\ \vdots \\ 0 \\ \vdots \\ v_{d_k}(n) \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad v_b(n) = \begin{pmatrix} v_{b_{11}}(n) \\ \vdots \\ v_{b_{1k}}(n) \\ \vdots \\ v_{b_{k1}}(n) \\ \vdots \\ v_{b_{kk}}(n) \end{pmatrix}, \quad v_\lambda(n) = \begin{pmatrix} v_{\lambda_{11}}(n) \\ \vdots \\ v_{\lambda_{1k}}(n) \\ \vdots \\ v_{\lambda_{k1}(n)} \\ \vdots \\ v_{\lambda_{kk}(n)} \end{pmatrix} \tag{3.58}$$

$$F_D = \begin{pmatrix} II_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & II_k \end{pmatrix}, \quad II_i = \begin{pmatrix} -1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{3.59}$$

$$H_d = \begin{pmatrix} 1 & 0 & \cdots & 0 & & & \mathbf{0} \\ \mathbf{0} & 1 & 0 & \cdots & 0 & & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \\ \mathbf{0} & \cdots & \mathbf{0} & 1 & 0 & \cdots & 0 \end{pmatrix}, \quad H_\lambda = \begin{pmatrix} u_1(n) & 0 & & \cdots & & 0 & \mathbf{0} \\ \mathbf{0} & & & & & & \mathbf{0} \\ \vdots & \mathbf{0} & u_2(n) & 0 & & \cdots & \\ \mathbf{0} & & & \cdots & & & \mathbf{0} \\ \vdots & \ddots & \ddots & & & \ddots & \vdots \\ \mathbf{0} & & & \cdots & & & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & u_k(n) \end{pmatrix} \tag{3.60}$$

$$H_b = \begin{pmatrix} u_1(n) & \cdots & u_k(n) & 0 & \cdots & 0 \\ \mathbf{0} & u_1(n) & \cdots & u_k(n) & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ \mathbf{0} & \cdots & \mathbf{0} & u_1(n) & \cdots & u_k(n) \end{pmatrix}. \tag{3.61}$$

3.6.1 Constraint Specification for Brand Substitution Component $g(n)$ for $k = 2$

$$\Gamma\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \Gamma\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} u_1^g \\ u_2^g \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{\bullet 1} & b_{\bullet 2} \end{pmatrix}\begin{pmatrix} u_1^g \\ u_2^g \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u_1^g \\ u_2^g \end{pmatrix}. \tag{3.62}$$

3.6.2 System Model for Brand Substitution Component on Simultaneous Equation for $k = 2$

$$\left(\begin{array}{c} b_{11}(n) \\ b_{\bullet 1}(n) \\ \hline b_{12}(n) \\ b_{\bullet 2}(n) \end{array}\right) = \left(\begin{array}{cc|cc} 1 & & & \\ & 0 & & \\ \hline & & 1 & \\ & & & 0 \end{array}\right)\left(\begin{array}{c} b_{11}(n-1) \\ b_{\bullet 1}(n-1) \\ \hline b_{12}(n-1) \\ b_{\bullet 2}(n-1) \end{array}\right) + \begin{pmatrix} v_{b_{11}}(n) \\ 0 \\ v_{b_{12}}(n) \\ 0 \end{pmatrix} \tag{3.63}$$

3.6.3 Constraint Specification for Category Expansion Component $z(n)$ for $k = 2$

$$\Gamma\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \Gamma\begin{pmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \end{pmatrix}\begin{pmatrix} u_1^g \\ u_2^g \end{pmatrix} = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{11} & \lambda_{22} \end{pmatrix}\begin{pmatrix} u_1^g \\ u_2^g \end{pmatrix}. \tag{3.64}$$

3.6.4 System Model for Category Expansion Component on Simultaneous Equation for $k = 2$

Due to the category expansion condition, the non-diagonal coefficients of $\lambda_{ij}(n)$ are zero, so that the following conditions hold:

$$\lambda_{\bullet 1}(n) = \lambda_{11}(n), \quad \lambda_{\bullet 2}(n) = \lambda_{12}(n). \tag{3.65}$$

This leads to the following system model:

$$\left(\begin{array}{c} \lambda_{11}(n) \\ \lambda_{\bullet 1}(n) \\ \hline \lambda_{22}(n) \\ \lambda_{\bullet 2}(n) \end{array}\right) = \left(\begin{array}{cc|cc} 1 & 0 & & \\ 1 & 0 & & \\ \hline & & 1 & 0 \\ & & 1 & 0 \end{array}\right) \left(\begin{array}{c} \lambda_{11}(n-1) \\ \lambda_{\bullet 1}(n-1) \\ \hline \lambda_{22}(n-1) \\ \lambda_{\bullet 2}(n-1) \end{array}\right) + \left(\begin{array}{c} v_{\lambda_{11}}(n) \\ v_{\lambda_{11}}(n) \\ v_{\lambda_{22}}(n) \\ v_{\lambda_{22}}(n) \end{array}\right). \tag{3.66}$$

3.7 Information Criterion – AIC

The information criterion, AIC, was developed by Akaike (1980a) to select the best parametric models among alternatives determined by the maximum likelihood method. A half of AIC value is an approximately unbiased estimate of the expected log-likelihood. AIC incorporates an approximate correction of the bias, involved in the maximized log-likelihood as an estimator of the average expected log-likelihood. The maximized log-likelihood is an information theoretic measure of the dissimilarity between two distributions, called as Kullback-Leibler information or the K-L number. The larger the measure, the greater the difference between the two distributions. The approximate bias reflected in AIC is equal to the number of parameters estimated in the model. The definition of AIC is given in the below:

$$\begin{aligned} \text{AIC}(m) &= -2(\text{maximized log likelihood of the model}) \\ &\quad +2(\text{number of estimated parameters in the model}) \\ &= -2\sum_{n=1}^{N} \log f_m(y_n|\hat{\theta}_m) + 2|\hat{\theta}_m|. \end{aligned} \tag{3.67}$$

In (3.67), $f_m(y_n|\hat{\theta}_m)$ denotes the likelihood and $|\hat{\theta}_m|$ denotes the dimension of the vector $\hat{\theta}_m$.

The AIC has proven to be extensively applicable in statistical data analysis and engineering modeling (see for example, Bozdogan (1994)). For the derivation of the AIC, refer to Kitagawa and Gersch (1996). A model with a smaller value of AIC is determined as a better model.

Chapter 4

4 Univariate Analysis of Scanner Sales

4.1 Determination of Price Function and Univariate Analysis of Category Sales

The methodology presented was applied to two-years daily milk data for the period of 1994/2/28–1996/3/3 (N=735) from two large super market stores. The data from one store, say store C, showed very little changes in brand prices, so that only time series analyses were conducted on the data (Out of 41 brands, one brand recorded different prices from the maximum for 16 days during the entire period, 14 days for another brand, 4 days for other two brands, no change at all for the rest of brands). Meanwhile, the data from the other store, store B, showed large variances on prices. Therefore, time series plus regression analyses were applied to the data from store B. Differences in sales movements of the two data sets are shown on Figure ?? as a surf graph, in which the sales of the first brand are recorded as the lowest line, the second lowest line is for the total sales of the first and the second brand, and the top line is for the category total. Similar to price movements, the graph on store C shows very small changes in its sales and brand share, having small variance of milk category total (3801). On the contrary, the graph on store B shows large changes in sales and the share as well as the large variance of the category total (19403).

Milk is a perishable product category that can be kept even in a cool place only less than a week, and newly delivered products are stocked on store shelves every day. Therefore, stockpile at households is not expected to occur often to a large extent.

Daily basis analysis is important for this kind of daily delivered category. There is no display activity and an advertisement in a flier is expected to be made not often. Because price promotion may be the only retail promotion for the category of milk, it can be seen as the major controllable factor by a retailer as sales promotion. Therefore,

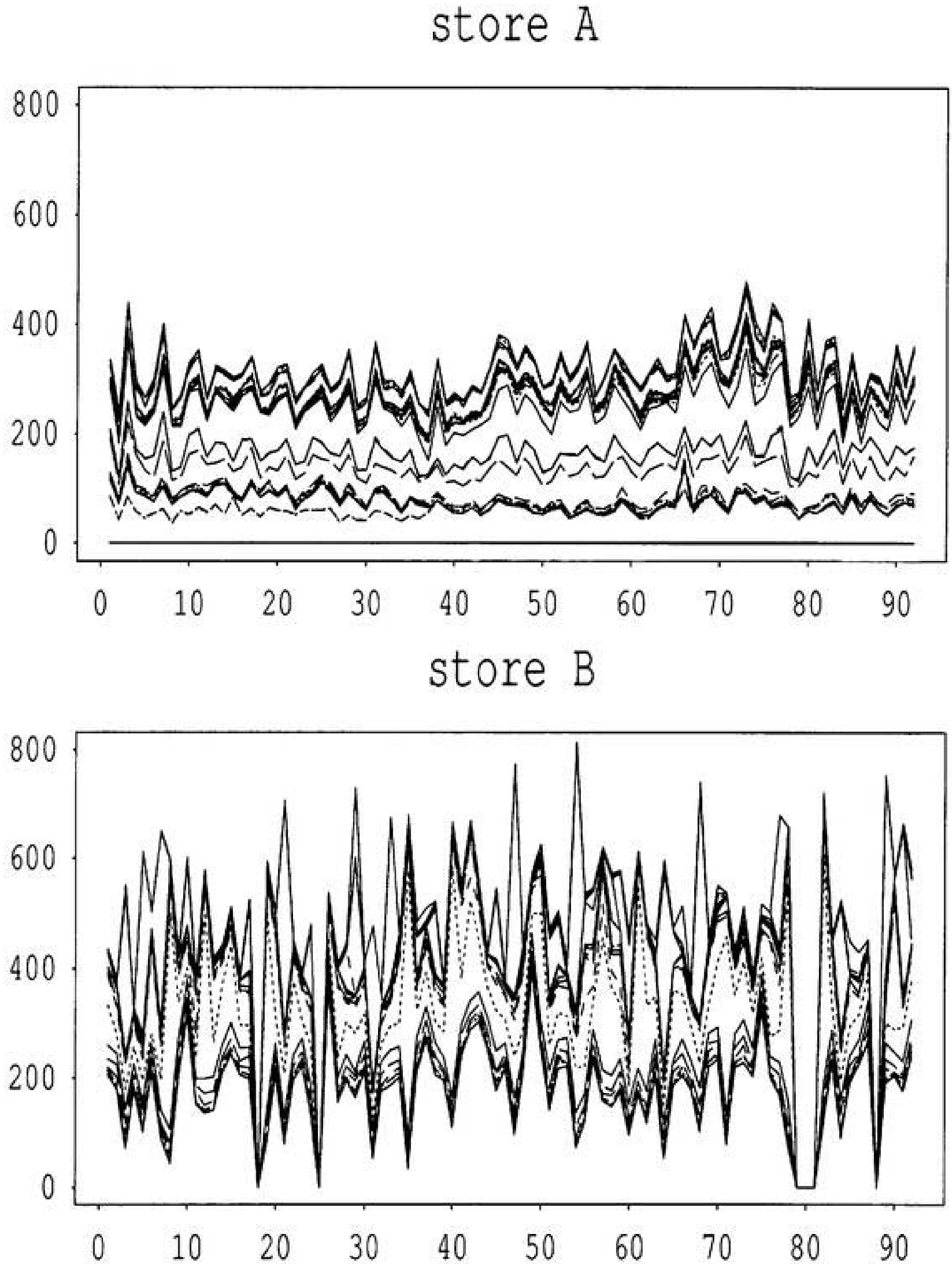


Figure 4.2: Litre sales of brand/category total on store A and store B for 92 days

price is chosen as the major source of explanatory variable on regression analysis. In this study, brand is defined as brand-size and sales quantity is defined as the litre of products purchased.

Time Series Analysis on Category Sales Data from Store A and Store B

The following time series models given in section 2.2 and 2.3 were fitted on category sales data from the both stores, respectively,

Model 1 (trend only)

$$y(n) = t(n) + w(n), \tag{4.68}$$

Model 2 (trend + cyclical)

$$y(n) = t(n) + d(n) + w(n). \tag{4.69}$$

A state space model is a recursive procedure and requires initial values of state mean and variance. If there exist some theoretically or empirically suggested values, use those values. Since we do not have the information on the initial values for our data, we determined them as follows: Initial value of state mean or variance for trend component was set to the mean or variance of the initial one fourth of the entire observations. The initial value of state mean for day-of-the-week component or for explanatory variable component was set to zero. The one of state variance for day-of-the-week component or for explanatory variable component was set to some large arbitrary value, which produces flatter prior distributions. If retailer managers are involved in the analysis of sales promotions and can access to various sources of information, use the empirical values as the initial values on specific product category. The specific knowledge on initial values will improve the result. There were 6 days of zero sales on milk category total for store C, 39 days for store B. All zero sales were treated as missing values. Those zero sales can be due to closed days of the stores or due to the disappearance of existed data. The maximum likelihood estimates of parameters were obtained by Broyden-Fletcher-Goldfarb-Shanno algorithm.

Summarized in Table 4.1 are the values of the log-likelihood, AIC, and the number of parameters. In Table 4.1, i denotes trend order and j denotes assumed period. For the both stores, the first order trend component Model (1,1), i.e., locally constant component, gave the best result among 3 polynomial models of Model 1 in equation (4.68). The result seems reasonable because data aggregation span is daily and very short and long-term component is supposed to change very little day to day.

Fixing the first order for the trend component, the number of days within a cyclical period were gradually increased for cyclical component. The result in Table 4.1 showed the Model (1,7) with 7 days in a cycle was selected as the best model for the both stores, confirming the existence of the day-of-the-week effect. Fixing 7 days in a cycle, we compared the order for the trend component, yet, the result was the same as before, choosing Model (1,7) as the best model for the both store.

Figure 4.3 shows the observed data, fitted trend, trend + day-of-the-week effect, and the residual component on the data from store C and store B, respectively. The residual component on store B shows considerably large variances compared with the one on store C, indicating the existence of the remaining effect by price promotion.

Regression Analysis on Each Brand and the Category Total on Store B Data

As seen in the above, "trend + day-of-the-week effect" model, Model (1,7), was found to be a better model than the "trend" model for the both data sets. The next step is to fit models having explanatory variable component as well. As described in section 2.4, the best price functions are to be determined among hypothesizable functions at first. Included here are examples of possible linear or non-linear regression models to show the

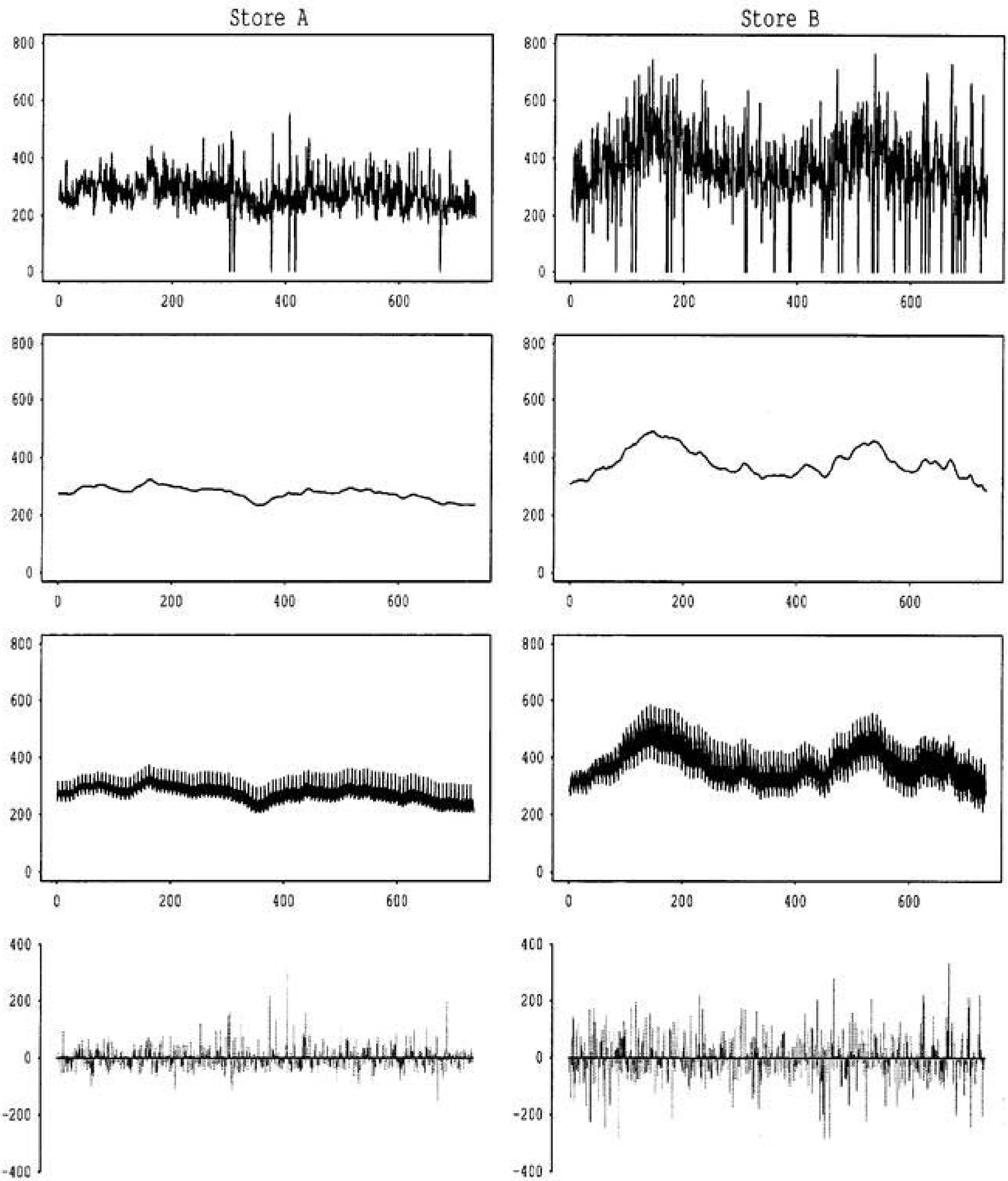


Figure 4.3: from top-to-bottom: the observation, fitted trend, trend + day-of-the-week effect, and residual component for store A and store B

Table 4.1: Log-likelihood and AIC for model (i, j) (i:trend order ; j:assumed period)

	store C		store B		
Model (i, j)	log-likelihood	AIC	log-likelihood	AIC	No. of Parameters
Model (1,1)	-3948	7903 †	-4223	8452 †	3
Model (2,1)	-3961	7931	-4233	8474	4
Model (3,1)	-3977	7964	-4250	8510	5
Model (1,2)	-3951	7913	-4226	8463	5
Model (1,3)	-3954	7920	-4229	8470	6
Model (1,4)	-3955	7923	-4231	8475	7
Model (1,5)	-3955	7927	-4233	8483	8
Model (1,6)	-3960	7939	-4235	8487	9
Model (1,7)	-3854	7728 ††	-4148	8315 ††	10
Model (2,7)	-3875	7771	-4166	8354	11
Model (3,7)	-3891	7807	-4183	8391	12
Model (1,8)	-3962	7946	-4236	8493	11
Model (1,9)	-3965	7955	-4240	8504	12
Model (1,10)	-3963	7953	-4241	8508	13

flexibility of our approach.

Brand Characteristics on Store B Data

Milk data from store B contain 46 brands, but the sales of major 4 brands, B4, B10, B11, and B30, account for 87% of the category sales. Therefore, prices of 4 brands were selected as the source of explanatory variable, and the sales of 4 brands and the category total, as 5 response variables. Although we are interested in only the relationship between category sales and the prices of 4 brands, the form of price function for each brand must be determined as the condition that intermittent pulse-like incremental sales are yielded by price reductions.

Figure 4.4 is the graphs on the data from store B in the period of 2 years. The first through the fourth graphs show movements of the major 4 brands (B4, B10, B11, B30). The last graph is for the sum of the 4 brands and shows a brand substitution mechanism in which an aggregation of 4 brands removed variances that existed in the original brand

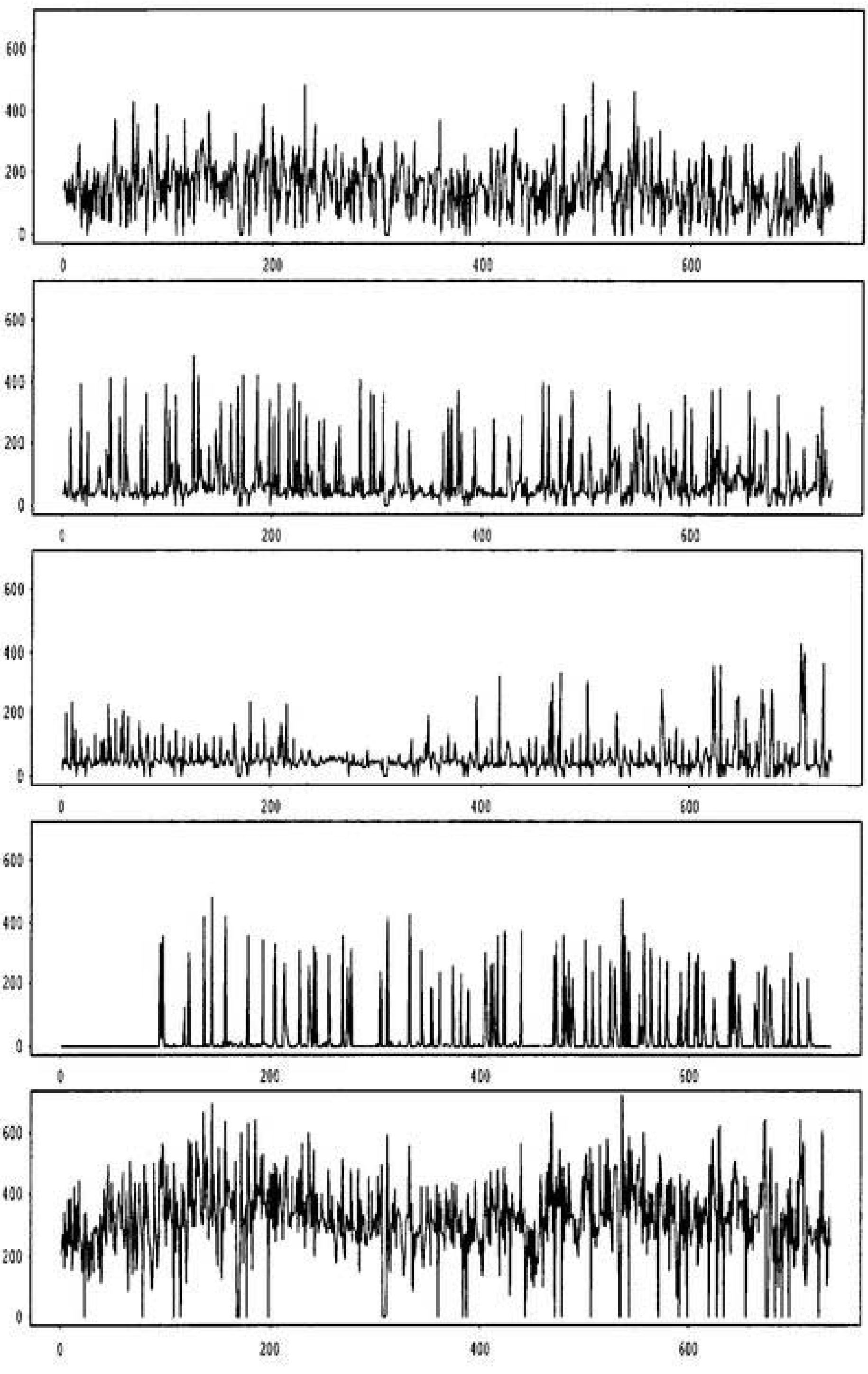


Figure 4.4: 4 Brands and the Sum of Daily Scanner Milk Data
Variance for 4 Brands and the Sum (7584; 7930; 3138; 8059; 17213)

data by the portion of brand substitution and exhibited an obvious seasonality.

Table 4.2 summarizes the maximum price during the periods for each brand, which is the substitute of the regular price. B10 is a national brand and has the highest regular (maximum) price. B11 is the low-fat type of the same brand of B10 and can be considered in a different segment from the other 3 brands. B4 is a private brand and has the lowest regular price among brands of regular type.

Table 4.2: The maximum price during the entire periods for each brand

brand	B4	B10	B11	B30
price	198	228	178	215

In Table 4.3, the competitive relationship between B4 and B10, or B4 and B30 can be recognized from the value of the cross correlations among the sales of each pair of 4 brands (-0.251 and -0.257).

Table 4.3: Cross correlations between each pair of 4 brands

	B4	B10	B11	B30
B4	1.000			
B10	-0.251	1.000		
B11	+0.060	-0.034	1.000	
B30	-0.257	-0.090	-0.039	1.000

Table 4.4 includes the correlations between the sales and the prices or the functions for each brand. No meaningful indications can be obtained from the positive correlations or almost zero correlation between the sales and the prices as shown on the second row of Table 4.4. On the other hand, high negative correlations, especially for B10 and B11, were recorded between the sales and the function f_1 shown on the third row, indicating the function f_1 is a good selection as a price function in general. On the fourth row to the ninth row of Table 4.4, the correlations between the sales and the function f_2 to the function f_7 were recorded by using the price functions obtained in the later analysis on

explanatory variable component.

Table 4.4: Correlations for each brand

brand	B4	B10	B11	B30
sales vs. prices	+0.324	-0.044	+0.004	+0.416
sales vs. f_1	-0.415	-0.714	-0.820	-0.571
sales vs. f_2	-0.483	-0.654	-0.755	NA
sales vs. f_3	-0.504 †	-0.669	-0.733	NA
sales vs. f_4	NA	NA	NA	-0.802
sales vs. f_5	-0.395	-0.201	-0.075	-0.727
sales vs. f_6	-0.504 †	-0.741 †	-0.842 †	-0.903 †
sales vs. f_7	-0.499	-0.720	-0.817	-0.807

The higher correlations on the third row of Table 4.4 implies that sales of each brand are affected by the depth of its price cuts as the primary cause. On the other hand, the lower correlations in Table 4.3, of which absolute values are less than 0.26, implies that the sales decline in its competitors due to own incremental sales is the secondary cause. Therefore, it is not unreasonable to assume that every consumer purchases a product independently by comparing among the prices of brands and price function is well described by differences in prices. There may be other factors which are confounded to price such as stock level. If stock level information is available, the inclusion as additional explanatory variable might yield a better result.

Figure 4.5 includes the graph for each brand on the sales (upper part) and price cuts (lower part) with days as x-axis, showing the pulse-like incremental sales occur rather coincidentally with the level of price cuts except B4.

Figure 4.6 shows scatter plots of the sales as x-axis against the price cuts (f_1) as y-axis for each brand, which are indicators on whether the sales can be well explained by its own price cut alone or not.

The graph on B4 shows that the sales corresponding to each level of price cut vary very much. This may be due to the reason that the sales of B4 were considerably influenced by its competitors of B10 and B30 as seen from negative correlations in Table 4.3 besides

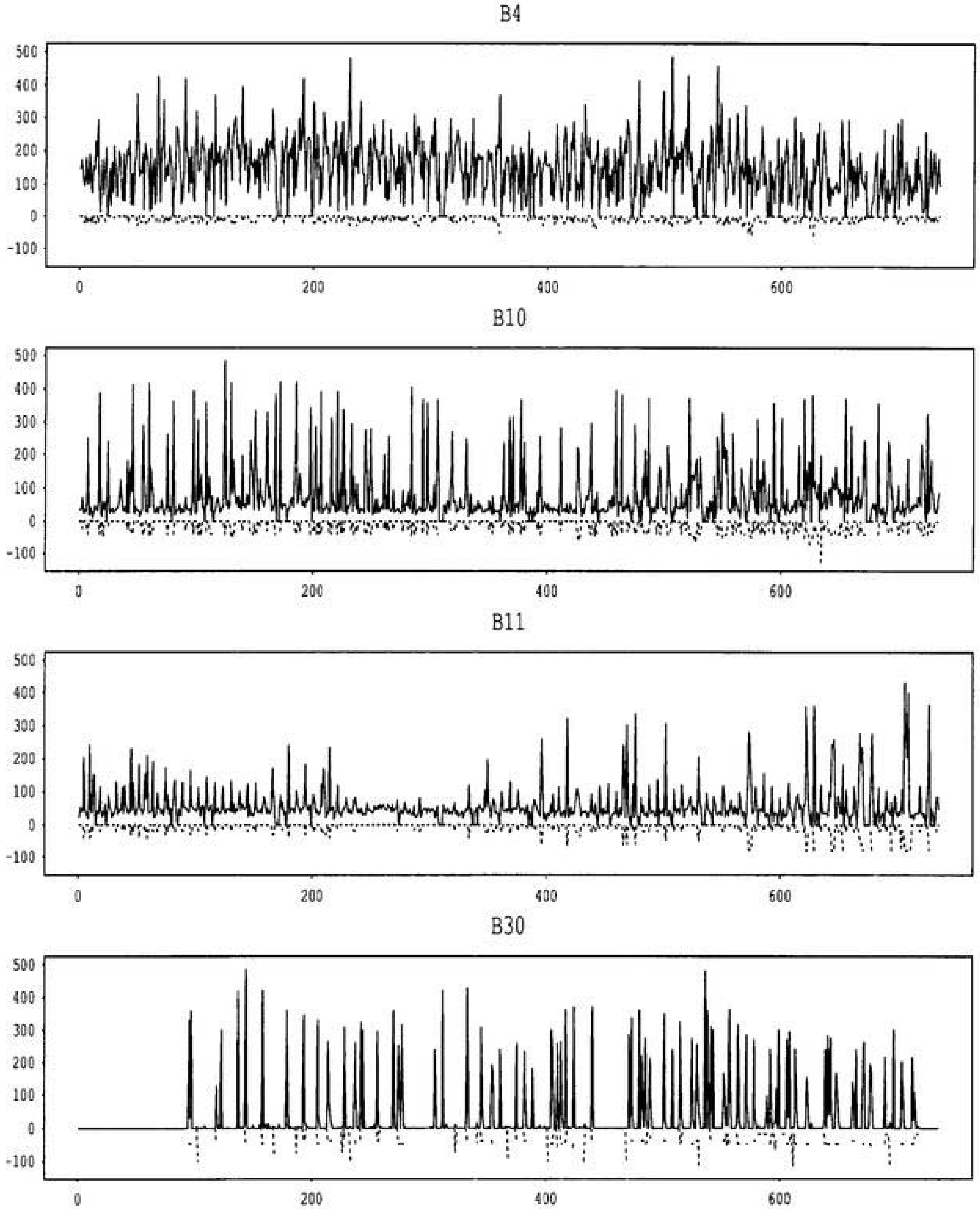


Figure 4.5: The movements of sales (upper part) and the price cuts (lower part) for each brand (B4, B10, B11, B30)

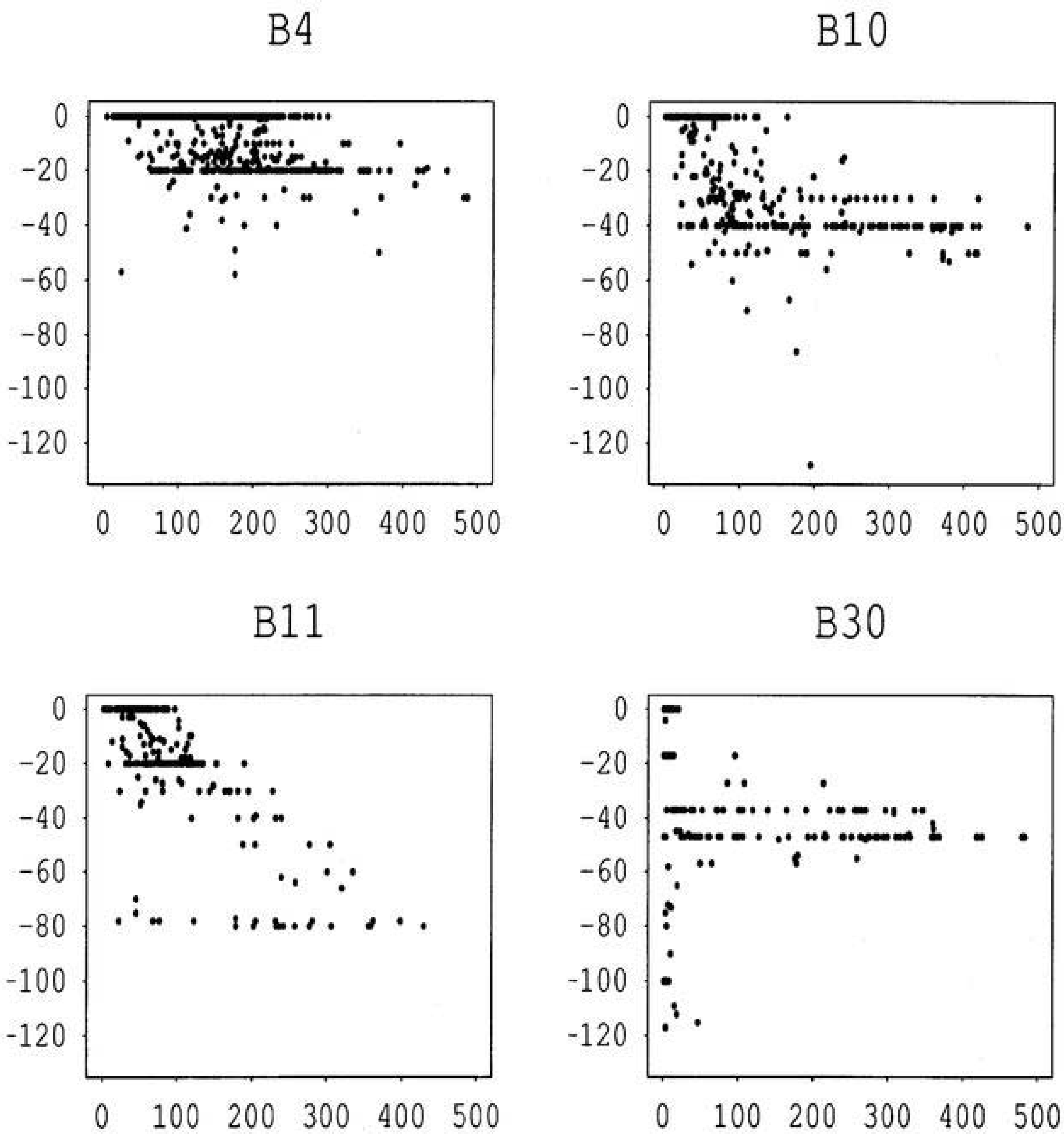


Figure 4.6: Scatter plots of the sales (x-axis) against the price cuts (y-axis) for each brand

by its own price cut.

The plots for B10 and B30 are roughly located in a triangle, indicating mixed effects from its own price cuts and from the competitors' sales and price cuts. In addition, B30 has considerable numbers of outliers near zero sales when the price cut is more than 60 yen.

The graph on B11 shows a clear linear relationship between sales and price cut level except plots with the price cut level of around 80 yen, where the sales seem to be distributed almost uniformly. This indicates that the sales of B11 can be well explained by its own price cut alone except the price cut around 80 yen.

Comparison of models with price function from f_1 to f_7 for 4 Brands

The models with the seven price functions together with the first order trend and the day-of-the-week effect were compared for 4 brands. The function f_1 accounts for the price cut level of a brand from the maximum price during the entire periods, in which only the relationship between the own price reductions and the sales is considered. For the price function f_2 and f_3, competitors' interference in the occurrence of incremental effects via its price cuts is considered. The function f_2 describes the condition that incremental effects are yielded only when the price cut is greater than that of its competitors. For the function f_3, incremental effects are yielded under the same condition as f_2, but the input level is determined by the difference of the price cut from that of its competitors. The function f_4 explains the condition that pulse-like incremental effects are yielded only when the price cut is within a certain range, having a lower and an upper threshold to eliminate the group that has a uniform distribution of price cut level when the sales were recorded as close to zero. The function f_5 is attributed to the condition that incremental effects are yielded only when the price is the lowest among competitors. The function f_6 is a non-linear version of the best selected model among f_1 through f_5, facilitating deal decay effects. The function f_7 is an exponential function of the best selected model among

f_1 through f_5, which has a saturation point in the level of inputs.

Analysis on B4

The correlations between sales and f_1 in Table ?? showed the lowest for B4. This may be due to influences by competitors regarding interference in the occurrence of incremental sales and the sales reduction because of competitors' price promotional incremental sales. We assumed that the prices of B4 can yield incremental sales only when there were no price cuts in other competitive brands. Our assumption is defined by the function f_2 or f_3 and is evaluated by comparing models with other price functions, together with the "trend" component and "trend + day-of-the-week" component. The function f_4 was not considered for B4, because plots in Figure 4.6 appeared to come from the same group.

The comparable models are summarized in Table 4.5, in which T denotes the existence of the first order trend component, W denotes the existence of the day-of-the week component, and f_i denotes the existence of the i th price function component. A blank denotes the non-existence of the component in that position. The model with price function f_3 was selected as the best model among them, supporting our assumption. The function f_3 could not have been obtained by weekly aggregated scanner data, because the combination of price cuts among competitors varys day to day. This competitive relationship of B4 with the other 3 brands may be described as "competitive interference". The deal decay effect was not recognized because the coefficient became close to zero, which makes the function f_6 virtually the same as the function f_3.

The result indicates comparative vulnerableness of B4 against B10, B11, and B30 in terms of producing incremental effects on price promotional occasions. This may come from weaker impacts of price reduction of B4 to consumers than other brands, of which regular price is the lowest among the regular type, or a reduced shelf space of B4 in competitors' promotional periods than otherwise.

Analysis on B10 and B11

Table 4.5: Log-likelihood and AIC of models on B4

Model(T,W,f_k)	log-likelihood	AIC	No. of Parameters
Model(T, ,)	-4032	8071	3
Model(T,W,)	-4013	8046	10
Model(T,W,f_1(B4))	-3954	7930	11
Model(T,W,f_2(B4))	-3920	7863	11
Model(T,W,f_3(B4))	-3909	7839 †	11
Model(T,W,f_5(B4))	-3958	7938	11
Model(T,W,f_6(B4))	-3909	7841	12
Model(T,W,f_7(B4))	-3915	7854	12

The price function f_2 and f_3 were superior to f_1 only for B4. For B10 and B11, price function f_2 and f_3 were inferior to f_1 as shown in Tables 4.6 and 4.7. This indicates that competitors price reductions do not interfere in yielding incremental effects for B10 and B11. The function f_4 was not considered for B10 and B11, because plots in Figure 4.6 appeared to come from the same group although there were some outliers.

The best selected price function for B10 was deal decay function, f_6, with a reset after 4 consecutive promotion runs ($\eta = 3$) and for B11, f_6 without reset. The main purpose of deal decay function is to include decaying effects of consumer's response to price promotions. The result for B10 shows that decaying effect is reset after 4 consecutive periods, indicating a possible situation that physically decreasing condition of stocked level was recovered by renewed stocks. There was an obvious upward (absolute value) outlier in price function, which is easily pinpoint as the plot having the price cut of more than 120 yen in Figure 4.6.

Analysis on B30 - Two Response Groups of Sales to Prices

The graph of B30 in Figure 4.6 showed that two groups have different movements against its price cuts. Although a linear relationship between price cuts and sales can be seen, there is a group that very large discount yields only negligible sales. Since milk is a perishable category, there is a practice that the prices of unsold products around two

Table 4.6: Log-likelihood and AIC of models on B10

Model(T,W,f_k)	log-likelihood	AIC	No. of Parameters
Model(T, ,)	-4121	8249	3
Model(T,W,)	-4118	8256	10
Model(T,W,f_1(B10))	-3886	7793 †	11
Model(T,W,f_2(B10))	-3940	7902	11
Model(T,W,f_3(B10))	-3928	7878	11
Model(T,W,f_5(B10))	-4112	8246	11
Model(T,W,f_6(B10))	-3864	7755 ††	13
Model(T,W,f_7(B10))	-3883	7790	12

Table 4.7: Log-likelihood and AIC of models on B11

Model(T,W,f_k)	log-likelihood	AIC	No. of Parameters
Model(T, ,)	-3697	7400	3
Model(T,W,)	-3684	7388	10
Model(T,W,f_1(B11))	-3318	6659 †	11
Model(T,W,f_2(B11))	-3390	6802	11
Model(T,W,f_3(B11))	-3429	6880	11
Model(T,W,f_5(B11))	-3684	7390	11
Model(T,W,f_6(B11))	-3289	6602 ††	12
Model(T,W,f_7(B11))	-3395	6815	12

days are lowered as much as by 30%. Therefore, it is very likely that those large price discount is to sell out left old products. The models on B30 with different lower and upper thresholds, i.e., the function f_4, were fitted by changing a threshold by yen. The price function f_2 and f_3 were not calculated because B30 has been stocked on the shelf intermittently, particularly on weekends, so that negative influence caused by its competitors price reduction was very limited. The model concerning the day-of-the-week-effect was not included due to regularly unstocked condition on weekdays. The price function, f_4 with the lower threshold of 18 and the upper threshold of 57 was selected as a better linear function than f_1. The best selected price function for B30 was deal decay function, f_6, without reset.

Table 4.8: Log-likelihood and AIC of models on B30

Model(T,W,f_k)	log-likelihood	AIC	No. of Parameters
Model(T, ,)	-1713	3431	3
Model(T, ,f_1(B30)	-1693	3394	4
Model(T, ,f_4(B30)	-1604	3220 †	6
Model(T, ,f_5(B30)	-1645	3298	4
Model(T, ,f_6(B30)	-1520	3054 ††	7
Model(T, ,f_7(B30)	-1606	3227	7

Figure 4.7 shows scatter plots of the sales as x-axis against the best selected function as y-axis for each brand, showing improvements in linear relationships from Figure 4.6.

Time Series + Regression Analysis on Category Sales Data from Store B

Time series plus regression analyses were conducted on Model 3 given in equation (4.70), by using the best price functions for the 4 brands determined in the above.

Model 3 (trend + day-of-the-week effect + explanatory variable effect)

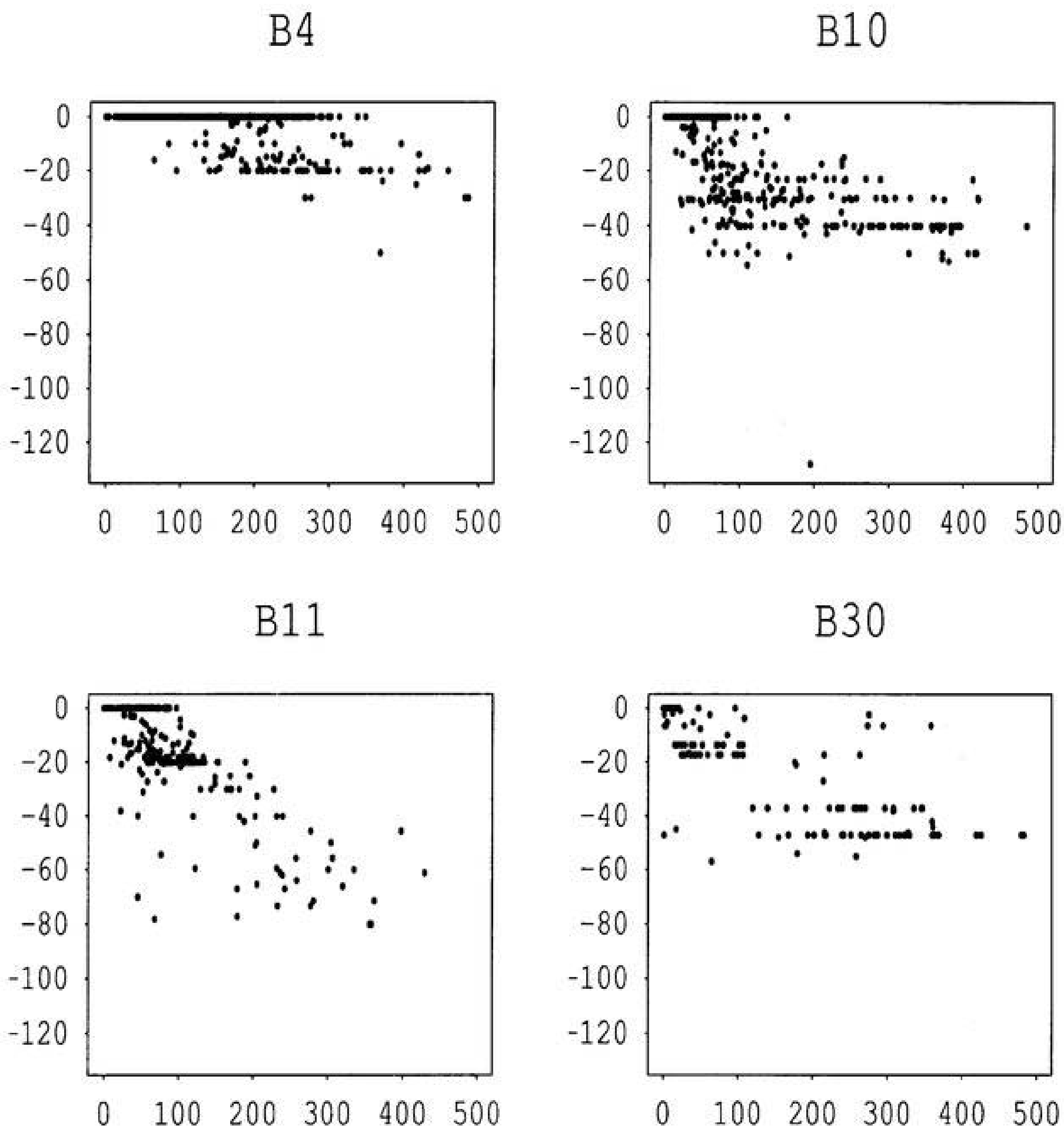


Figure 4.7: Scatter plots of the sales (x-axis) against the best selected price function (y-axis) for each brand

$$y(n) = t(n) + d(n) + x(n) + w(n) \tag{4.70}$$

Table 4.9 includes the values of the log-likelihood, AIC, and the number of parameters. The mark * (on B10) shows an involvement of an extra function to treat an outlier for B10. The 12th model with 3 price functions of B10*, B11, and B30 has the lowest AIC, but violates the condition in equation (??). The 11th model with the same 3 price functions that have the same coefficient gave the best result. This means that the price cuts on B10, B11, and B30 were effective to increase both its own brand sales and the category sales, indicating oppositely that the price function of B4 has an incremental effect only on the brand sales, but not on the category sales. Namely, the incremental sales of B4 were only yielded at the expense of competitors.

Table 4.9: Log-likelihood and AIC for models on category sales

No	Model(T,W,f_k)	log-likelihood	AIC	No. of Parameters
1	Model(T, ,)	-4223	8452	3
2	Model(T,W,)	-4148	8315	10
3	Model(T,W,f_3(B4))	-4152	8326	11
4	Model(T,W,f_6(B10))	-4146	8318	13
5	Model(T,W,f_6(B10*))	-4144	8316	14
6	Model(T,W,f_6(B11))	-4109	8242	12
7	Model(T,W,f_6(B30))	-4122	8272	14
8	Model(T,W,f_6(B10*,B11))	-4106	8243	16
9	Model(T,W,f_6(B10*,B30))	-4108	8252	18
10	Model(T,W,f_6(B11,B30))	-4076	8185	16
11	Model(T,W,f_6(B10*,B11,B30))	-4064	8164 †	18
12	Model(T,W,f_6(B10*,B11,B30))	-4061	8163	20

Deal decay function was better than linear function for 3 brands out of four (B10, B11, and B30). For the other brand, B4, although deal decay was not recognized, instead, the brand seems to be interfered by simultaneously promoted competitive brands in yielding incremental effect. The result is interesting and instructive, indicating the following hypothesis:

When incremental effects are produced by price promotions, some competitive relations exist either between competitive promotional occasions of brands or between the first promotional occasion of a brand and the succeeding promotional occasions.

Figure 4.8 shows the comparison between the best fitted model from Model 2 and the one from Model 3 on the store B data. From the top to the bottom, they are the observed data, the fitted trend, trend + day-of-the-week effect, trend + day-of-the-week effect + explanatory variable effect, and residual component, respectively. A comparison between the third graph from the top on Model 2 and the one on Model 3 shows that the explanatory variable effect was included in the day-of-the-week effect on Model 2. This means that the price promotion appeared to be carried out on particular days within a week, e.g. on week-ends, and the price promotional effects on the particular days are counted as the day-of-the-week effect. Therefore, if a two-step model fitting procedure were conducted, the day-of-the-week effect would be over-evaluated as in Model 2 and would not give appropriate results. On Model 3, the day-of-the-week effect and the price promotional effects are separated. The smaller residual in Model 3 than that on Model 2 shows improvements in modeling.

Figure 4.9 gives a comparison on analytical results between store A and store B and reveals how price promotion "intervene" normal movements of the sales. If price promotions were not carried out, the sales for store B data would have been explained only by trend and day-of-the-week effect as do for the sales from store A. The additional component of explanatory variable on store B accounts for remaining pulse-like effects after trend and day-of-the-week effect being removed. The level of the day-of-the-week effect on Store B data appeared to be reasonable compared with that on Store A data.

Figure 4.10 shows how incremental effects on the category sales decays for each level of price cuts on B10, B11, and B30, fixing the other explanatory variable being zero. These values will be useful in estimating how much incremental category sales are gained

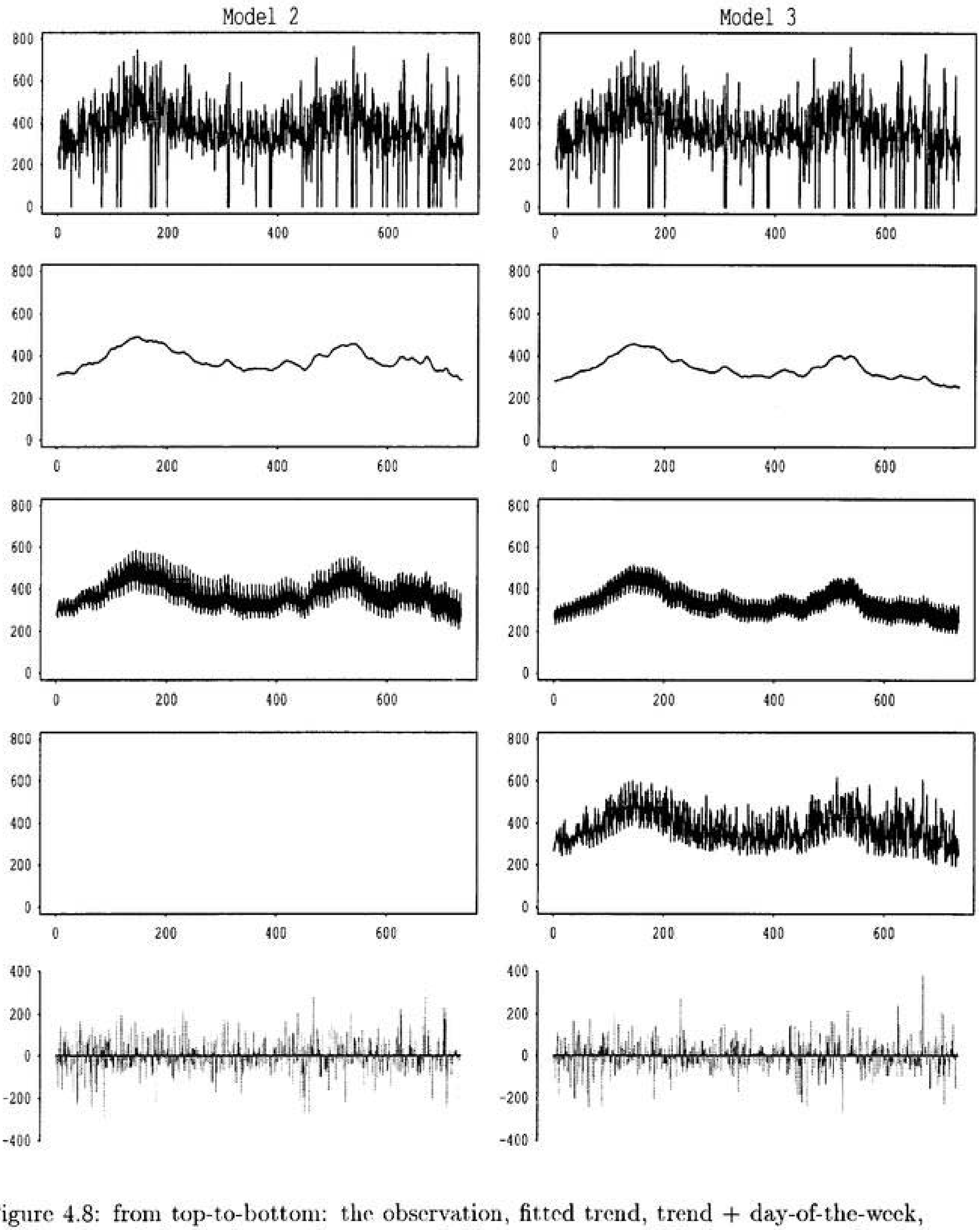


Figure 4.8: from top-to-bottom: the observation, fitted trend, trend + day-of-the-week, trend + day-of-the-week + explanatory variable component, residual component, for the best Model 2 and the best Model 3 on store B data

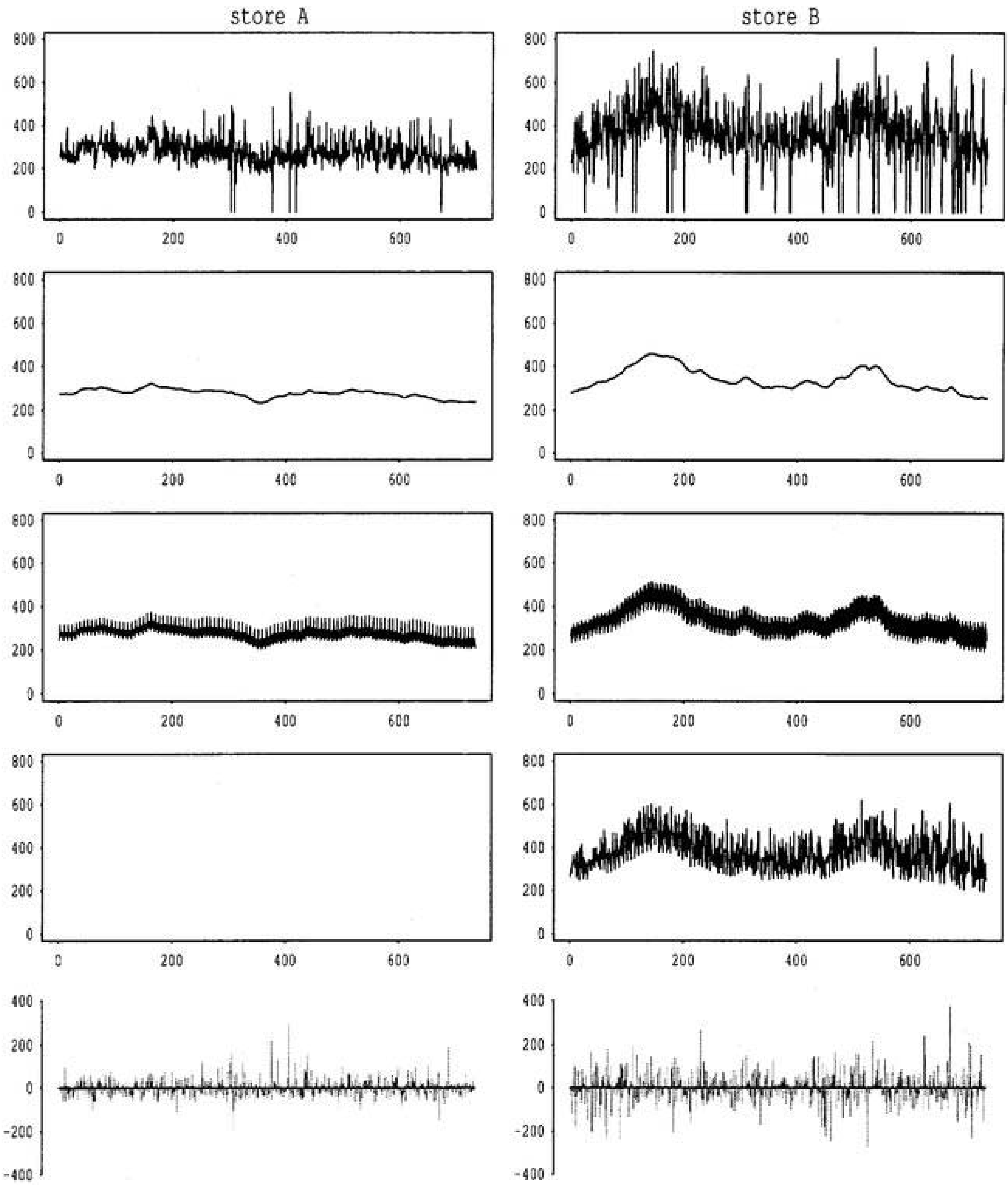


Figure 4.9: from top-to-bottom: the observation, fitted trend, trend + day-of-the-week, trend + day-of-the-week + explanatory variable component, residual component, for store A and store B

by changing the price of one brand at a time, and how much incremental sales for each brand decays as time passes when successive promotion runs are continued.

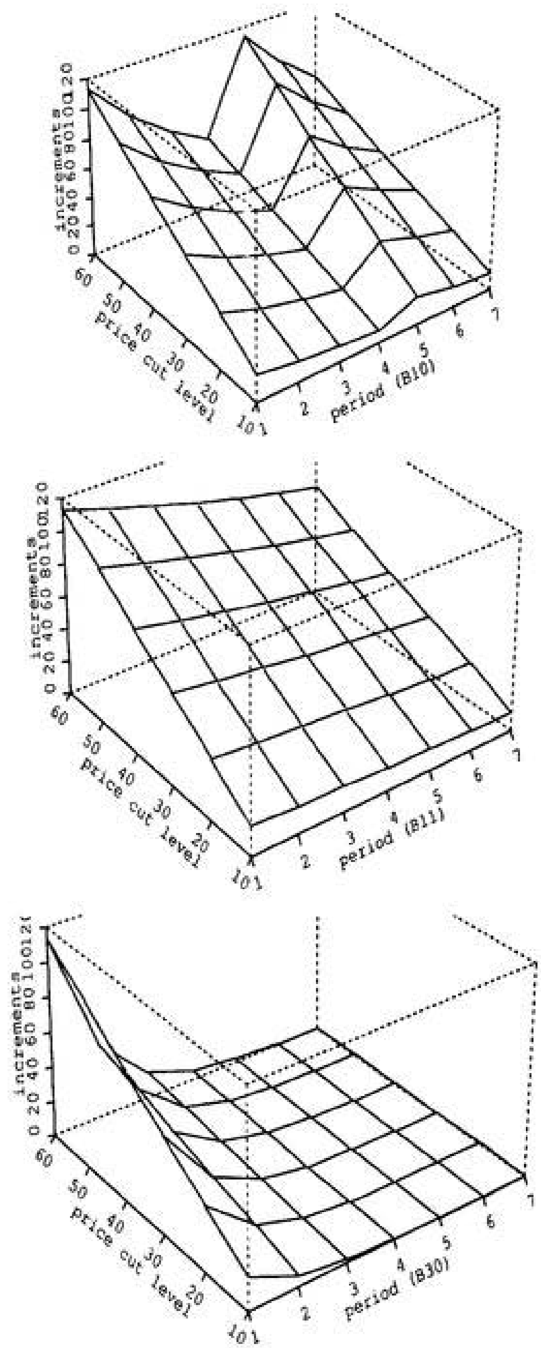


Figure 4.10: Decay of incremental sales (vertical axis) for each price cut level on B10, B11, and B30

Chapter 5

5 Multivariate Analysis on Scanner Sales

The presented vector state space models with time-varying regression parameters were applied to 3 scanner data sets of milk category. One is brand sales and prices of weekly aggregated data for the period of 1994/1st week (Jan.3) – 1995/10th week (Mar.12) (N=62) from store A, and another, those of 2 years daily aggregated data for the period of 1994/2/28 – 1996/3/3 (N=735) from store B. The third is only brand sales of the same daily basis of 2 years period from store C, in which price reduction did not occur so often. The data sets from store B and store C are the same as in Chapter 4. All of the three stores are large super market stores. The top 4 brands and the others total were analyzed for the data from the three stores. In this study, brand is defined as brand-size and sales quantity is defined as the litre of products purchased.

The left-hand side plots of Figure 5.11 are for the weekly observations from store A, which show mildly changing phenomena in incremental sales as well as in baseline sales. The left-hand side plots of Figure 5.12 are for the daily observations from store B, also show mild changes in incremental sales, and there appear to exist more variation in daily data than in weekly observations possibly due to day-of-the-week effect (Kondo and Kitagawa (1998)). The left-hand side plots of Figure 5.13 are the data from store C, in which mild changes are detected only in baseline sales. Therefore, considered models were such that environmental factors including seasonality or brand competitions affect both baseline sales and incremental sales.

5.1 Time Series Analysis on Data through Store A to Store C

The model 1 given in equation (5.71) was fitted to both weekly and daily data from the three stores. The model 2 that considers also day-of-the-week component given in equation (5.72) was applied only to daily basis data from store B and store C.

Model 1 (baseline only)

$$y(n) = t(n) + w(n), \tag{5.71}$$

Model 2 (baseline + cyclical)

$$y(n) = t(n) + d(n) + w(n). \tag{5.72}$$

A state space model has sequential estimation procedures and requires initial values of state mean and variance. The one of state variance for day-of-the-week component or for explanatory variable component was set to some large arbitrary value, which produces flatter prior distributions. A discussion on the determination of initial values is in Harvey (1989, P120 – 125).

All zero sales were treated as missing values. Those zero sales can be due to closed days of the stores or due to the disappearance of existed data. For data from store A, there were 6 weeks of modified observations out of 62 due to missing daily data or problems in quality, so that those data are treated as missing observations. The maximum likelihood estimates of parameters were obtained by a quasi-Newton numerical optimization procedure based on Broyden-Fletcher-Goldfarb-Shanno algorithm.

The values of the log-likelihood, AIC, and the number of parameters on the analysis of Model 1 were summarized in Table 5.10 where i denotes trend order and j denotes assumed period. The baseline component for brand A4 of store A and brand B4 of store B were fixed because the both are newly entered brands, show very little movements in baseline sales, and have fewer observations than other brands. Among 3 polynomial models of Model 1 in equation (5.71), the first order trend component Model (1,1), i.e., locally constant component, gave the best result for all of the stores. The result indicated that data aggregation span of daily or weekly basis is so short that long-term component was represented by the first order stochastic equation, a locally constant component.

For the data from store B and store C, fixing the first order for the trend component, the number of days within a cyclical period were gradually increased for cyclical component. The day-of-the-week component was not included for brand B4 of store B because B4 are rather in regularly unstocked conditions on weekdays. The results in Table 5.10 showed the Model (1,7) with 7 days in a cycle was selected as the best model for the both stores, confirming the existence of the day-of-the-week effect for multivariate data as well. Fixing 7 days in a cycle, we compared the order for the trend component, yet, the result was the same as before, choosing Model (1,7) as the best model for the both store. These results correspond to our earlier work.

The right-hand side plots of Figure 5.11 show fitted baseline for store A on brands A1 – A4 and the others total. The right-hand side plots of Figure 5.12 and Figure 5.13 show the observations and fitted baseline + day-of-the-week effect for store B and store C for the brands, respectively. For the store C data, the residual is expected to be relatively small except some occasional spikes of incremental sales. For the data from store A and B, it can be easily seen the existence of large variances in the residual. Further analyses on price promotions are conducted in the next section.

Table 5.10: Log-likelihood and AIC for model (i, j) (i:trend order ; j:assumed period)

	store A		store B		store C		
Model (i, j)	log-L	AIC	log-L	AIC	log-L	AIC	No. of Parameters
Model (1,1)	-1997.3	4044.7	-19653.0	39356.0 †	-18897.4	37844.7	25
Model (2,1)	-2042.9	4145.8	-19862.3	39784.6	-19274.2	38608.4	30
Model (3,1)	-2094.8	4259.6	-23665.2	47400.4	-24768.5	49607.0	35
Model (1,2)			-19661.4	39392.8	-18826.7	37723.4	35
Model (1,3)			-19669.5	39419.0	-18837.5	37755.0	40
Model (1,4)	NA		-19671.1	39432.2	-18845.1	37780.2	45
Model (1,5)			-19681.7	39463.4	-18851.8	37803.6	50
Model (1,6)			-19684.6	39479.2	-18859.4	37828.8	55
Model (1,7)			-19507.0	39134.0 ††	-18740.0	37600.3	60
Model (2,7)	NA		-19727.2	39584.4	-19181.3	38492.6	65
Model (3,7)			-23973.6	48087.2	-25183.1	50506.2	70

5.2 Price Function on Each Brand and Category Total for Store A Data

Brand Characteristics on Store A Data

Out of 28 brands, the sales of the top 4 brands, A1 through A4, account for 75% of the category sales. Prices of 4 brands were selected as the source of explanatory variable, and the sales of 4 brands and the category total, as 5 response variables.

Table 5.11 summarizes the maximum price during the periods for each brand, which is the substitute of the regular price. A1 is a national brand and has the highest regular (maximum) price. A2 is the low-fat type of the same brand of A1 and can be considered in a different segment from the other 3 brands. A3 is a private brand and has the lowest regular price among brands of regular type. A4 is a new entry into the market with the same maximum price as that of A1.

Table 5.11: The maximum price during the entire periods for each brand

brand	A1	A2	A3	A4
price	228	188	195	228

In Table 5.12, a strong competitive relationship between A1 and A3 and a weak relationship between A1 and A4 can be recognized from the value of the cross correlations among the sales of each pair of 4 brands (-0.800 and -0.276).

Price Function

As described in section 2.4, the models with four price functions together with the first order baseline component were compared for 4 brands and functions were determined at first among the functions, f_1 - f_3, and f_6 are illustrated in Figure (to be included).

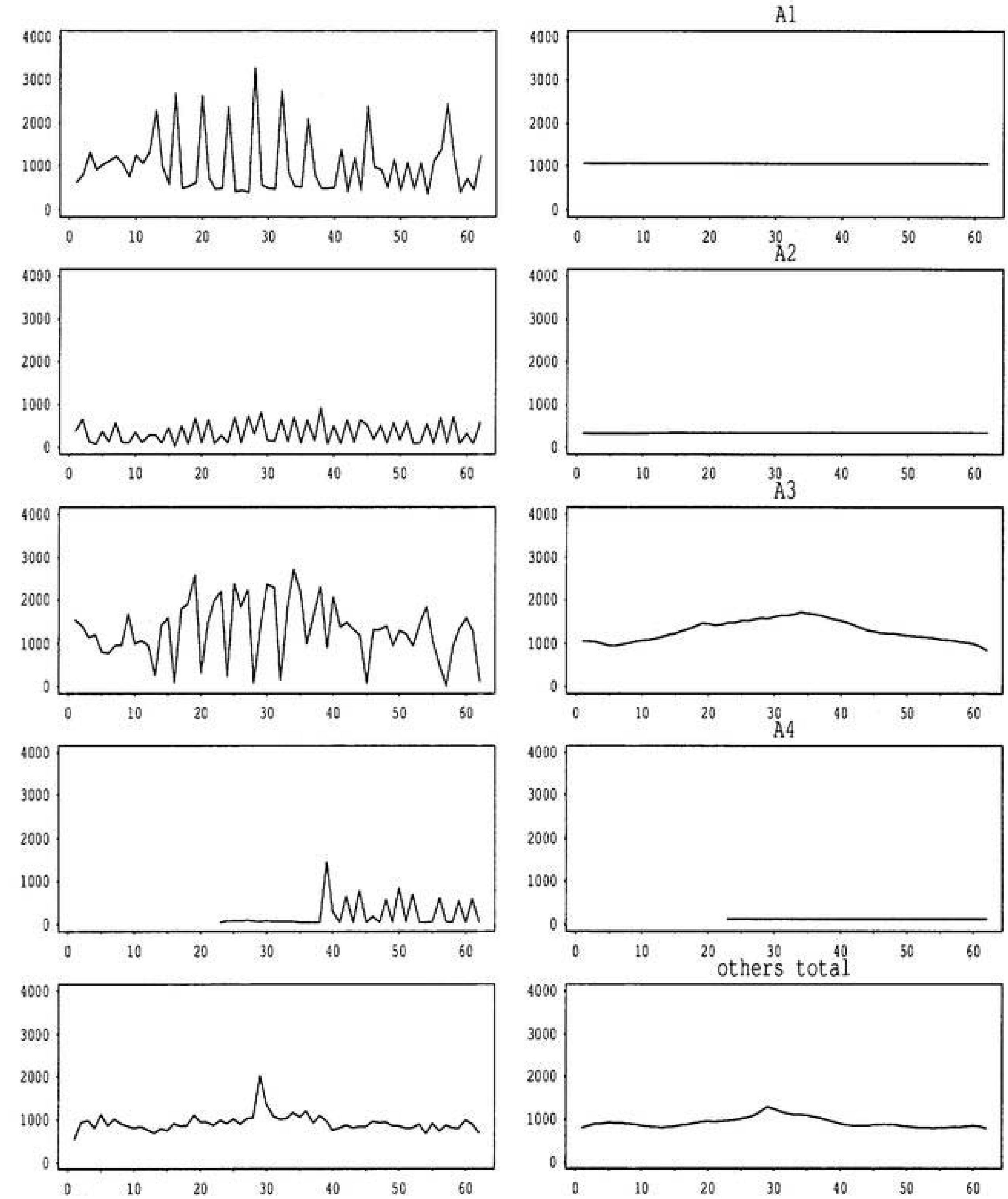


Figure 5.11: Observations (left); fitted trend (right) for store A data (1994/week01 - 1996/week10)

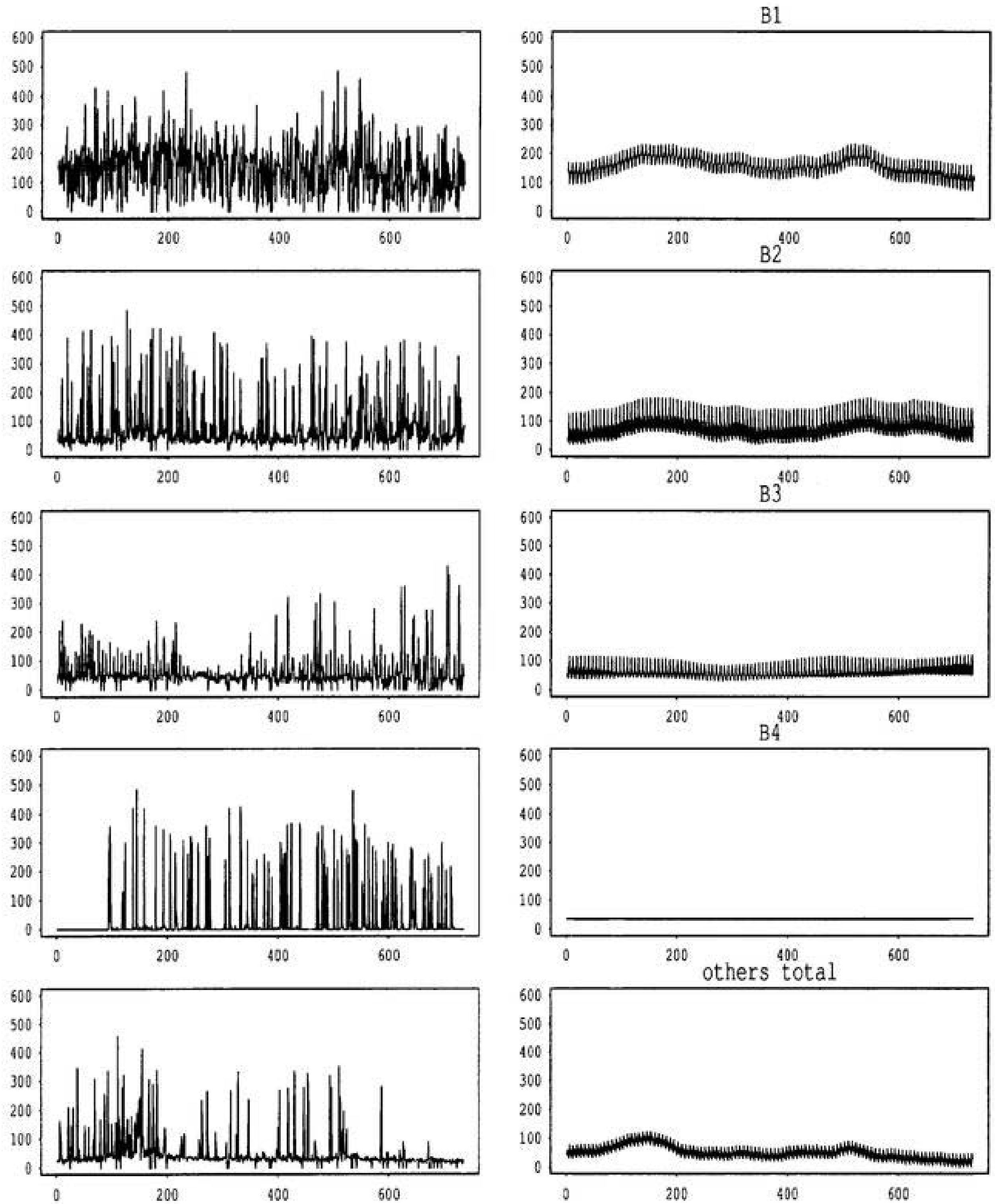


Figure 5.12: Observations (left); fitted trend + day-of-the-week component (right) for store B data (1994/2/28 - 1996/3/3)

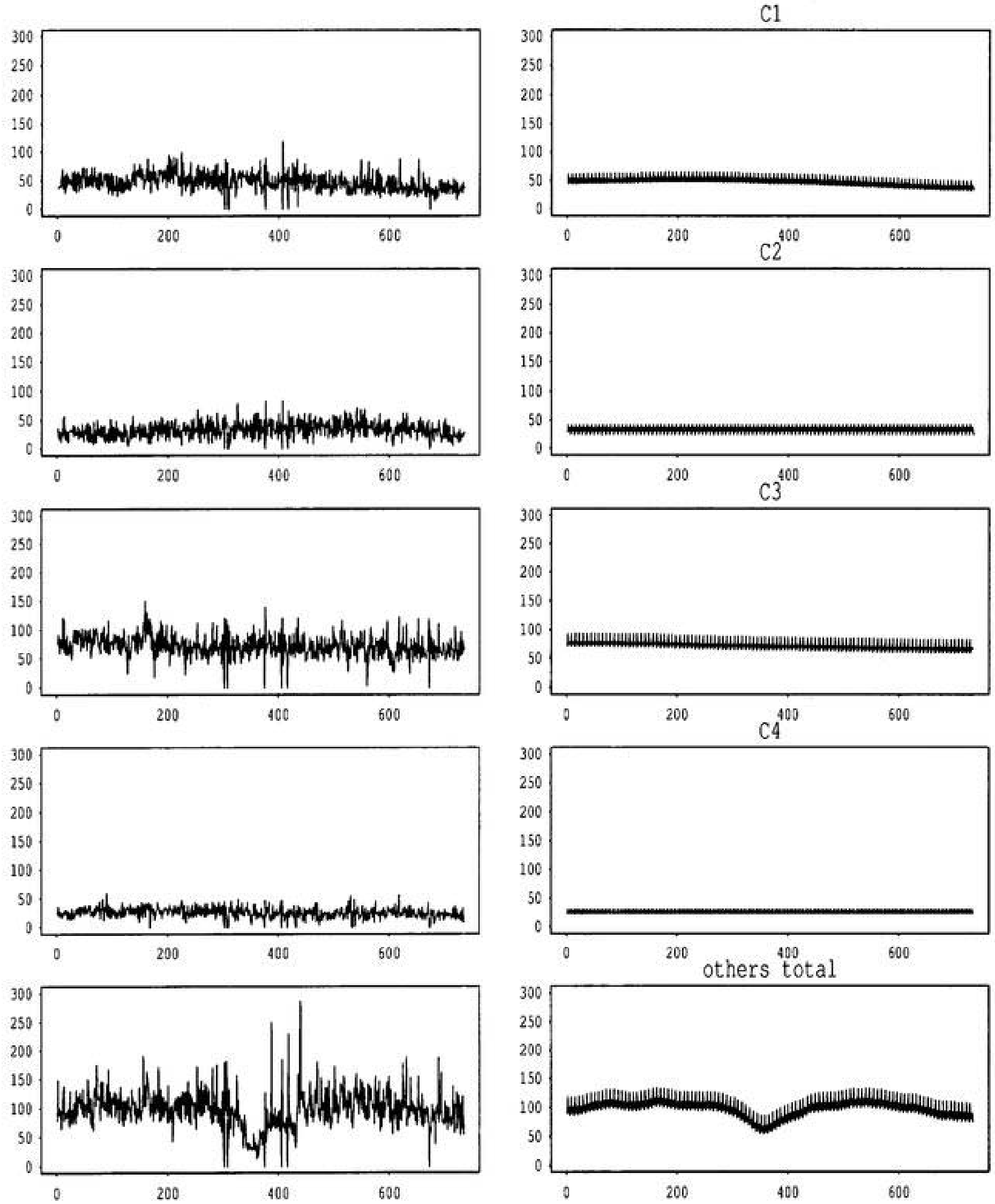


Figure 5.13: Observations (left); fitted trend + component (right) for store C data (1994/2/28 - 1996/3/3)

Table 5.12: Cross correlations between each pair of 4 brands

	A1	A2	A3	A4
A1	1.000			
A2	-0.094	1.000		
A3	-0.800	0.145	1.000	
A4	-0.276	-0.083	-0.076	1.000

The functions, f_4, f_5, f_7 were not included among the alternatives. This is because the function f_5 did not yield any better results than the simplest function f_1, similarly, no better for the function f_7 compared with any simpler instantaneous functions in any cases in Chapter 4. The function f_4 is to select a range which contribute to increase the sales of a brand, but there appeared to have no clear cut range for store A data.

Price Function on A3

The results on comparable models are summarized in Table 5.13, in which T denotes the existence of the first order baseline component, and $f_i(x)$ denotes the existence of the i th price function component of brand x. A blank denotes the non-existence of the component in that position.

The model without any price function was selected as the best model among the alternatives. In our previous work on daily data from store B in Chapter 4, the price cut of a private brand was effective to increase at least the own brand sales, in which the model with price function f_3 was selected as the best model. For this weekly aggregated scanner data from store A, however, none of the models with price function was fitted well to the data.

Price Function on A1, A2, and A4

The results on comparable models are summarized in Tables 5.14, 5.15, 5.16. Among the instantaneous price function, the simplest price function f_1 was the best choice for the brands, A1 and A2. The price function f_2 was superior to f_1 only for brand A4. The best selected price function for A1 and A4 was deal decay function, f_6, with a reset after

Table 5.13: Log-likelihood and AIC of models on A3

Model(T,f_k)	log-likelihood	AIC	No. of Parameters
Model(T,)	-496.2	998.5 †	3
Model(T,f_1(A3))	-500.2	1008.3	4
Model(T,f_2(A3))	-500.0	1008.0	4
Model(T,f_3(A3))	-497.6	1003.3	4
Model(T,f_6(A3))	-497.6	1007.3	6

2 consecutive promotion runs ($\eta = 1$). The main purpose of deal decay function is to include decaying effects of consumer's response to price promotions. The result for the two brands shows that decaying effect is reset after two consecutive periods, indicating that consumers' responses to the discount decline in the next week of the first promotion run and recovers two weeks later. For A2, the log-likelihood was improved with the deal decay function, but the AIC did not become the minimum because that the price reduction itself did actually follow almost the same decaying pattern as explained in the above.

Figure (to be inserted) shows how incremental effects on the category sales decay for each level of price cuts on A1 and A4, fixing the other explanatory variable being zero. These values will be useful in estimating how much incremental category sales are gained by changing the price of one brand at a time, and how much incremental sales for each brand decays as time passes when successive promotion runs are continued.

Table 5.14: Log-likelihood and AIC of models on A1

Model(T,f_k)	log-likelihood	AIC	No. of Parameters
Model(T,)	-454.1	914.1	3
Model(T,f_1(A1))	-439.9	887.8 †	4
Model(T,f_2(A1))	-445.5	899.0	4
Model(T,f_3(A1))	-451.0	910.0	4
Model(T,f_6(A1))	-436.2	884.3 ††	6

Table 5.15: Log-likelihood and AIC of models on A2

Model(T, f_k)	log-likelihood	AIC	No. of Parameters
Model(T,)	-389.5	785.0	3
Model(T, f_1(A2))	-364.2	736.4 †	4
Model(T, f_2(A2))	-378.9	765.8	4
Model(T, f_3(A2))	-382.9	773.7	4
Model(T, f_6(A2))	-363.5	738.9	6

Table 5.16: Log-likelihood and AIC of models on A4

Model(T, f_k)	log-likelihood	AIC	No. of Parameters
Model(T,)	-233.5755	473.1510	3
Model(T, f_1(A4))	-225.6	459.2	4
Model(T, f_2(A4))	-219.3	446.6 †	4
Model(T, f_3(A4))	-230.3	468.6	4
Model(T, f_6(A4))	-212.4	436.7 ††	6

Price Function on Category Sales Data from Store A

Table 5.17 includes the values of the log-likelihood, AIC, and the number of parameters. The model with price function f_1(A2) of a low-fat type brand gave the best result. This means that only the price cuts on A2 contribute to increase the category sales as well as the brand sales. The price cuts on A1 and A4 were effective to increase its own brand sales only and not the category sales.

Table 5.17: Log-likelihood and AIC for models on category sales

No	Model(T, f_k)	log-likelihood	AIC	No. of Parameters
1	Model(T,)	-464.5	935.1	3
2	Model(T, f_6(A1))	-468.9	949.7	6
3	Model(T, f_1(A2))	-461.7	931.5 †	4
4	Model(T, f_6(A4))	-468.7	946.4	6

ismspc5

5.3 Time Series plus Regression Analysis on Data from Store A and Store B

Time series plus (constant parameter) regression analyses without day-of-the-week effect were conducted on Model 3 given in equation (5.73) for the store A data and on Model 4 given in equation (5.74) for data from store B. The best price functions determined in the above is used for the store A data and the ones determined in Chapter 4 was used for the store B data.

Model 3 (trend + explanatory variable effect)

$$y(n) = t(n) + x(n) + w(n) \tag{5.73}$$

Model 4 (trend + day-of-the-week effect + explanatory variable effect)

$$y(n) = t(n) + d(n) + x(n) + w(n). \tag{5.74}$$

Tables 5.18 and 5.20 summarizes the values of the log-likelihood, AIC, and the number of parameters on Model 3 for store A data and on Model 4 for store B data, respectively, with constant explanatory variable parameter ($\sigma_b^2 = 0$) in equation 2.28. Here, the letter of T, W, X denotes the existence of baseline component, day-of-the-week component, and explanatory variable component. "Backward" variable selection was performed with the reducion order of a variable that yields the smallest AIC when it was removed from the full model. The process is continued until the AIC happens to increase. Then, the variable is returned to the subset, the variable of the next order is removed, and the calculation of AIC is performed. The procedure is continued until the last regression parameter is removed.

Among the subset models, the model with the smallest AIC value, model M12 for store A data and model M2 for store B data, was chosen. The next step was "forward" variable selection by adding a variable that yields the smallest AIC when it was added to the chosen subset. However, for data sets of store A or store B, no variable was added because the addition of any variable to the subset did not decrease the AIC value, although the addition of a certain variable slightly improved the log-likelihood for store B data (Model M+12).

The best selected constant regression parameter models reflect how each explanatory variable of price function affects each sales of brands as seen in Tables 5.19 and 5.21. The variables having the number in a parenthesis were included in the best selected model. Table 5.19 shows the results on store A data. Except A3, price reduction for each brand contribute to increase the own sales, which corresponds to the results when price function was determined. The sales of private brand A3 were negatively influenced by the price cuts of competitor, A1, to a very large extent. The sales of national brand, A1, were negatively affected by price cuts of competitor, A4. The sales of low-fat-type, A2, those of A4, and the ones of the others total of 24 brands were not affected at all by either of price cuts of competitive 4 brands. Table 5.21 shows the result on store B data that price reduction for each brand contribute to increase at least the own sales. The sales of private brand B1 were negatively influenced by the price cuts of competitors, B2 and B4, the sales of national brand, B2, by the price cut of competitor's brand, B4. The sales of low fat type brand B3, those of B4, and the sales of the others total of 42 brands were not affected at all by either of price cuts of competitive 4 brands.

5.4 Comparison of Constant and Time-varying Regression Coefficient Estimates

Time-varying Parameter model was established by specifying $\sigma_b^2 \neq 0$, in the selected constant parameter model (Model M12 for store A and Model M2 for store B). A comparison was made between the constant parameter models and the time-varying parameter models in Tables 5.22 and 5.23 where the letter of T, W, X are the same as before and the suffix C and M denotes constant parameters and time-varying parameters. The results show the models with time-varying parameters was selected as the best model.

Table 5.18: Log-likelihood and AIC for Model 3 on Store A Data

Model (T, X)	log-likelihood	AIC	No. of Parameters
Model (T,)	-1997.3	4024.7	25
Model Full	-1979.3	4088.6	65
Model M10	-1975.2	4078.4	64
Model M17	-1971.0	4064.1	61
Model M4	-1966.9	4049.7	58
Model M9	-1959.2	4032.3	57
Model M5	-1959.2	4026.3	54
Model M20	-1955.4	4012.8	51
Model M14	-1951.8	4003.5	50
Model M6	-1950.7	3999.3	49
Model M18	-1950.1	3992.2	46
Model M15	-1949.3	3988.6	45
Model M13	-1946.1	3980.3	44
Model M8	-1943.6	3973.4	43
Model M2	-1942.4	3964.8	40
Model M11	-1940.9	3959.9	39
Model M12	-1938.7	3953.5 †	38
Model M16	-1940.8	3951.6	35
Model M3	-1951.8	3967.7	32
Model M1	-1966.8	3991.7	29
Model M19	-1966.3	3984.7	26
Model M7	-1997.3	4024.7	25

Table 5.19: Sales vs. Price function for Store A data

Sales	Price function			
	U_{A1}	U_{A2}	U_{A3}	U_{A4}
Y_{A1}	(1)	6	11	(16)
Y_{A2}	2	(7)	12	17
Y_{A3}	(3)	8	13	18
Y_{A4}	4	9	14	(19)
Y_{OT}	5	10	15	20

Table 5.20: Log-likelihood and AIC on Constant Parameter Model for Store B data

Model (T, W, X)	log-likelihood	AIC	Parameters
Model (T,W,)	-19507.0	39134.0 †	60
Model Full	-18293.2	36806.4	110
Model M4, 9,14	-18276.0	36758.0	103
Model M8	-18270.9	36739.8	99
Model M20	-18267.8	36727.6	96
Model M18	-18260.9	36707.8	93
Model M12	-18260.0	36704.0	92
Model M11	-18253.8	36689.6	91
Model M5	-18251.4	36680.8	89
Model M10	-18248.8	36667.6	85
Model M15	-18246.4	36660.8	84
Model M3	-18240.3	36644.6	82
Model M2	-18237.6	36635.2 ††	80
Model M17	-18308.2	36770.4	77
Model M16	-18308.2	36764.4	74
Model M6	-18340.0	36820.0	70
Model M1	-18451.6	37039.2	68
Model M7	-18808.8	37745.6	64
Model M13	-19262.2	38650.4	63
Model M19	-19507.0	39134.0	60
Model M+12	-18237.0	36636.0	81

Table 5.21: Sales vs. Price function for Store B data

Sales	Price function			
	U_{B1}	U_{B2}	U_{B3}	U_{B4}
Y_{B1}	(1)	(6)	11	(16)
Y_{B2}	2	(7)	12	(17)
Y_{B3}	3	8	(13)	18
Y_{B4}	4	9	14	(19)
Y_{OT}	5	10	15	20

The inclusion of just one common parameter σ_b^2 in system noise variance given in equation (2.28) could improve the results for the data sets from the both store. Moreover, smooth movements of regression parameters were achieved by the weights of system noise variance in baseline component, otherwise over-fitting could have easily occurred for the data with sharp pulse-like spikes.

Figure 5.14 (store A) and Figure 5.15 (store B) show the movements of time-varying parameters of the baseline component (left) and the ones of incremental component (right) influenced by the own price function except A3, which is influenced by the competitors price function. The both lines in the left and the right of Figure 5.14 and Figure 5.15 show roughly reversing images in its movements for each brands, except the movements for A1 in Figure 5.14 and the ones for B2 in Figure 5.15. This is the most interesting finding in the analysis that baseline component and incremental component are often affected by environmental factor very similarly, but sometimes differently. The parameter movements of B2 in baseline component appeared to reflect maily seasonality, while those in incremental component, competitive relations. On the other hand, the reverse can be seen in A1. This may indicate that when strong competitive movement is dominant, background movements of seasonality becomes difficult to be visible. For brand A3, the movements of time-varying parameter is the one influenced by the competitors' price reduction, so that the sign is opposite to other brands. The extent of parameter variation is large in the order of A3 (private brand) or A1((national brand), and then A2 (low-fat type), and A4 for store A, and B1 (private brand), B2 (national brand), B3 (low-fat type), and B4 for store B.

The top to the bottom of Figure 5.16, Figure 5.17 and the right-handside of Figure 5.18 show the observations, the fitted trend, trend + day-of-the-week effect, trend + day-of-the-week effect + explanatory variable effect, and residual component for store B data, respectively. It can be easily seen that the ways of parameter changes are different among brands. The results gave a clear contrast among brands in terms of whether the parameters in baseline or in incremental are prone to environmental changes such as seasonality or brand competition. The parameters in baseline and in incremental moves fairly well for the private brand of B1 whose sales are affected by two competitors' price discounts. This can be easily seen in Figure 5.18, which compares fixed parameter model both in baseline and incremental (left-handside) with the best fitted time-varying parameter model (right-handside) on store B, for private brands, B1. For store A data, the same comparisons were made on brand A1 in Figure 5.19. The bottom of the Figures show the seasonality remains in residual component for a constant parameter model. This means that, for private brands, a time-varying regression parameter model is more appropriate than a constant regression parameter model (or OLS procedure). An importance should be placed on finding what kinds of brands show large movements in the parameter to determine whether time-varying parameter model is required or not.

The accumulation of these kinds of results by detailed data analysis by vector state space model may provide a basis to give a solution concerning the dispute over the condition that naive time series model give better results than the constant regression parameter models for market share forecasting (Brodie and de Kluyver (1987)) or visa versa (Kumar and Heath (1990)).

Table 5.22: Log-likelihood and AIC for Constant and time-varying Parameter Model for store A data

Model (T, X)	log-likelihood	AIC	No. of Parameters
Model (T_C,)	-2024.5	4135.1	20
Model (T_M,)	-1997.3	4024.7	25
Model M12 (T_C, X_C)	-1954.8	3959.7	25
Model M12 (T_M, X_C)	-1938.7	3937.4	30
Model M12 (T_M, X_M)	-1933.9	3959.8 †	31

Table 5.23: Log-likelihood and AIC for Constant and Time-varying Parameter Model for store B data

Model (T, W, X)	log-likelihood	AIC	No. of Parameters
Model $(T_C,,)$	-19747.1	39534.2	20
Model $(T_M,,)$	-19653.0	39356.0	25
Model $(T_C, W_C,)$	-19637.9	39385.8	55
Model $(T_M, W_M,)$	-19507.0†	39134.0	60
Model M2 (T_C, W_C, X_C)	-18458.9	37041.8	62
Model M2 (T_M, W_M, X_C)	-18237.6	36609.2	67
Model M2 (T_M, W_M, X_M)	-18224.5 ††	36585.1	68

Finally, a comparison between the second plots of the right-handside in Figure 5.12 on Model 2 and the third plots in Figure 5.16 (right) on Model 4 shows that the explanatory variable effect was included in the day-of-the-week effect on Model 2. This means that the price promotion appeared to be carried out on particular days within a week, e.g. on week-ends, and the price promotional effects on the particular days are counted as the day-of-the-week effect. Therefore, if a two-step model fitting procedure were conducted, the day-of-the-week effect would be over-evaluated as in Model 2 and would not give appropriate results. On Model 4, the day-of-the-week effect and the price promotional effects are separated.

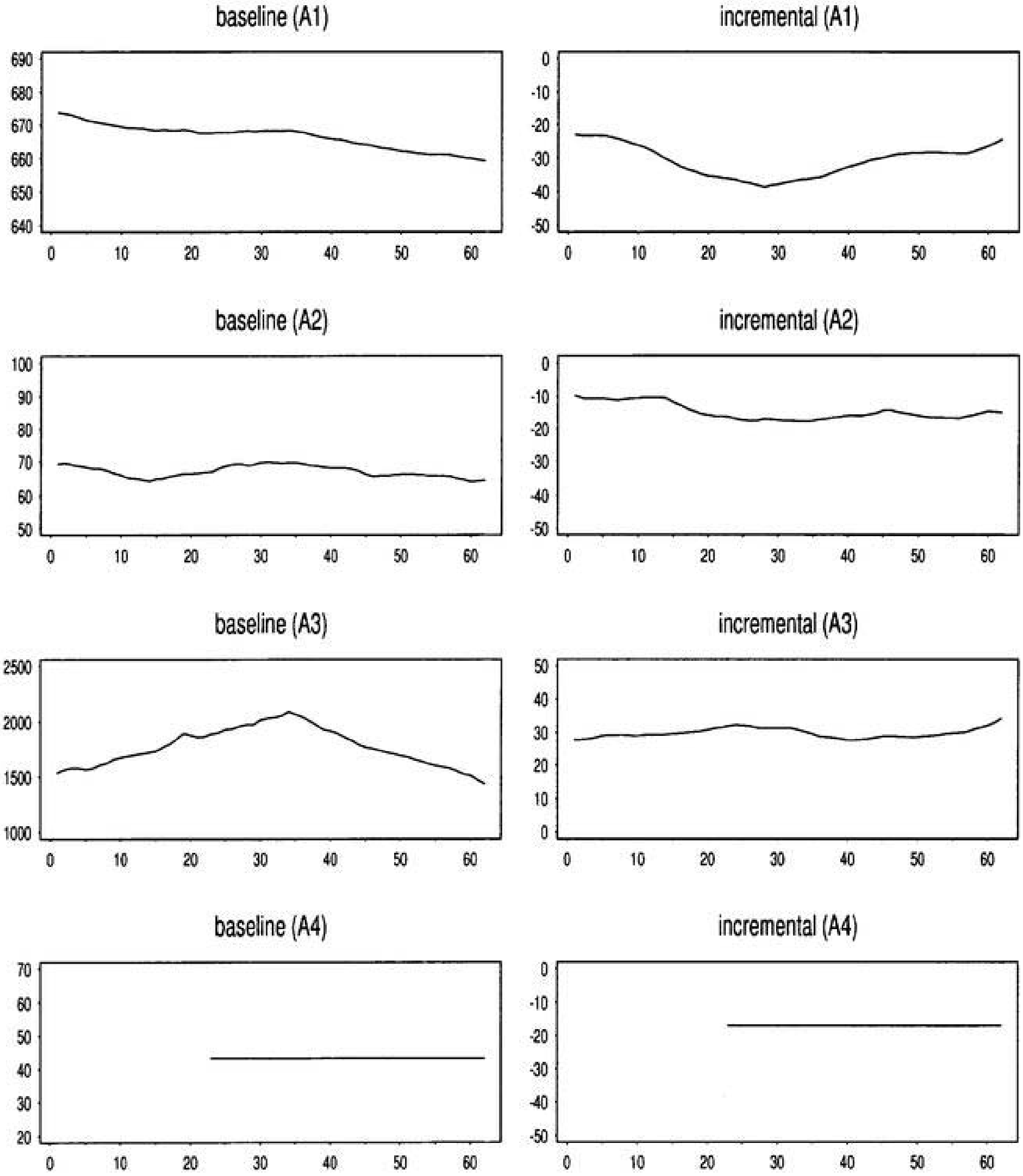


Figure 5.14: time-varying parameter movements – the baseline component (left-handside) and incremental component (right-handside) for store A data

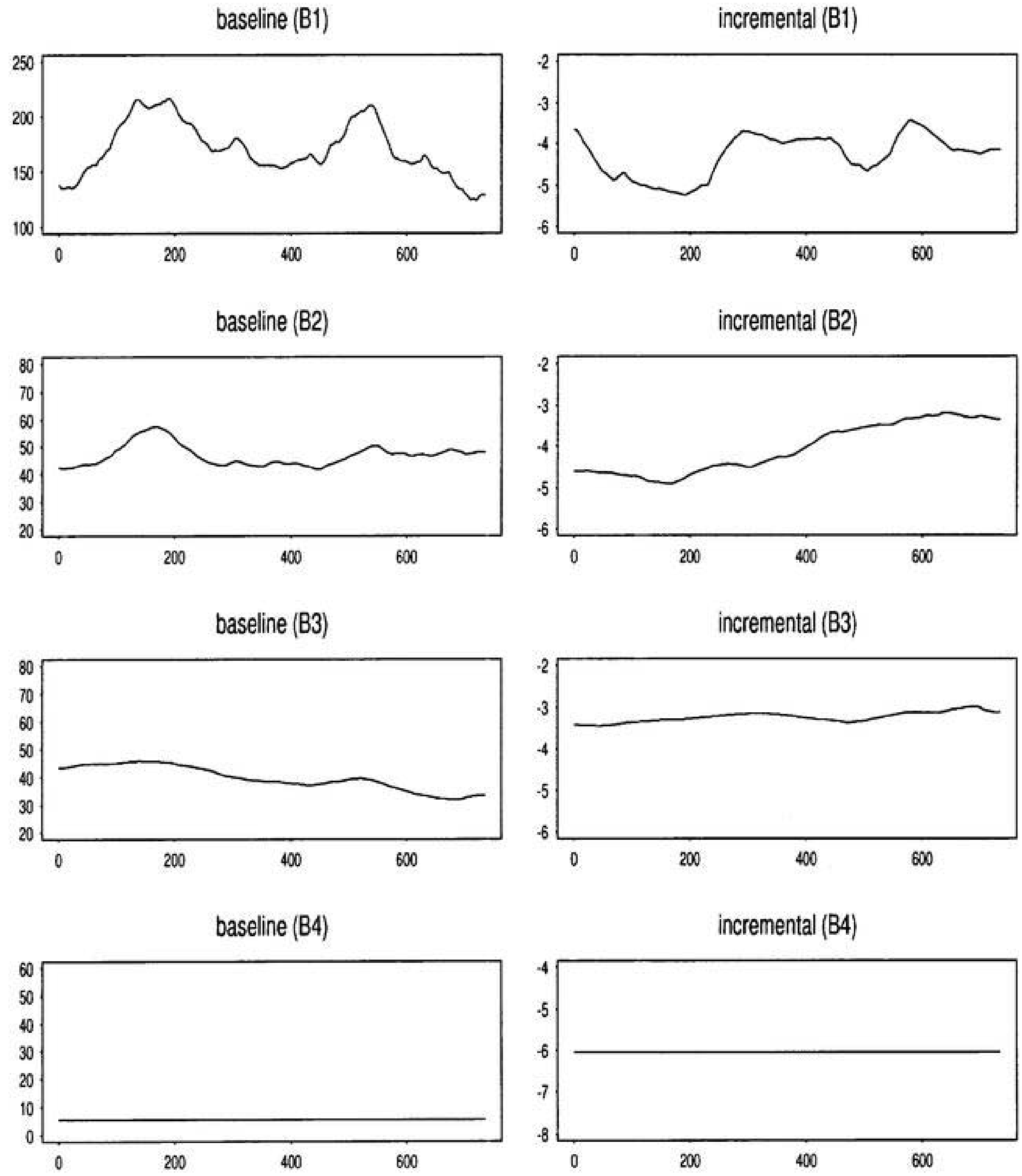


Figure 5.15: time-varying parameter movements – the baseline component (left-handside) and incremental component (right-handside) for store B data

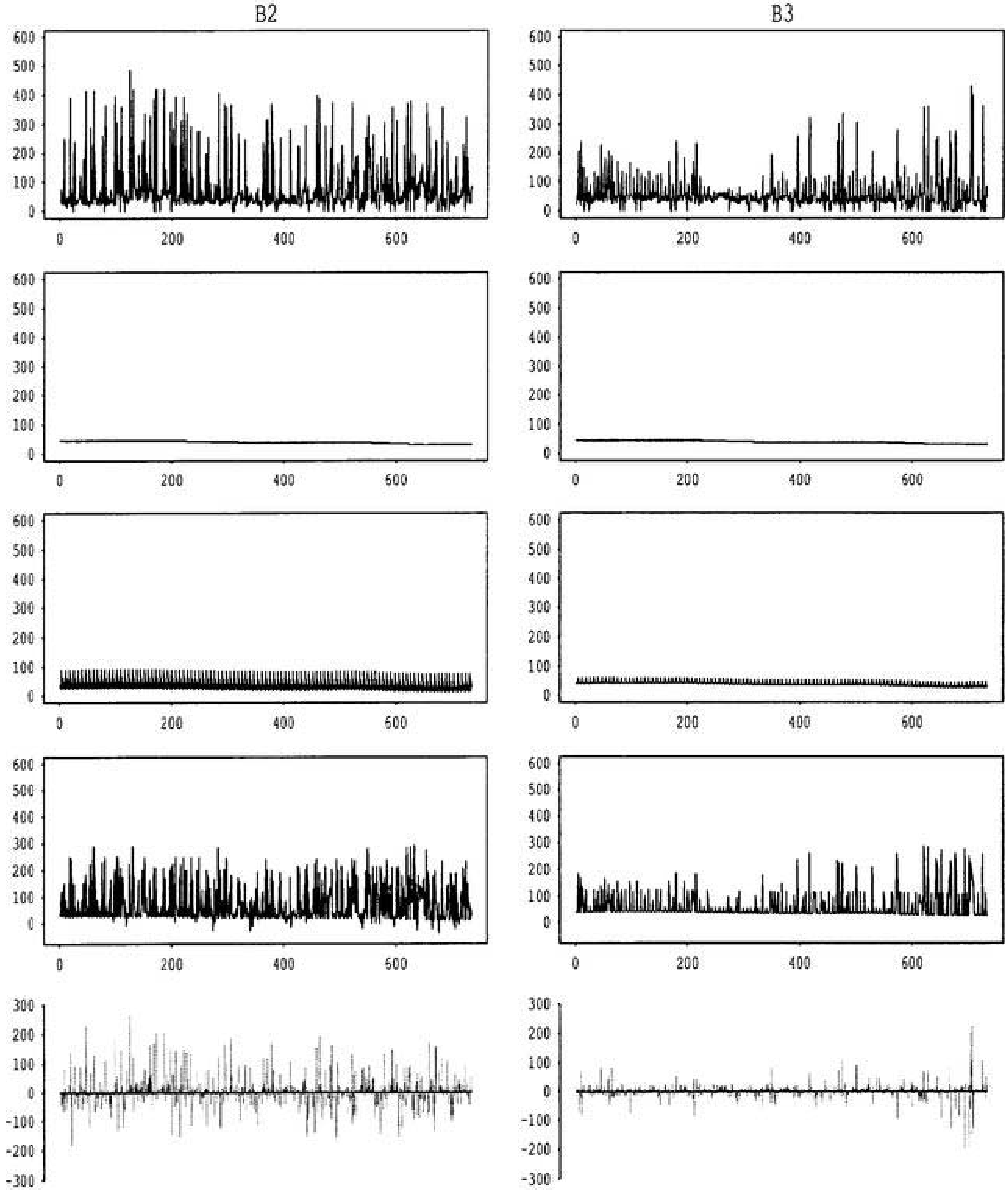


Figure 5.16: from top-to-bottom: the observation, fitted trend, trend + day-of-the-week, trend + day-of-the-week + time-varying explanatory variable component, residual component for store B

(1994/2/28 - 1996/3/3)

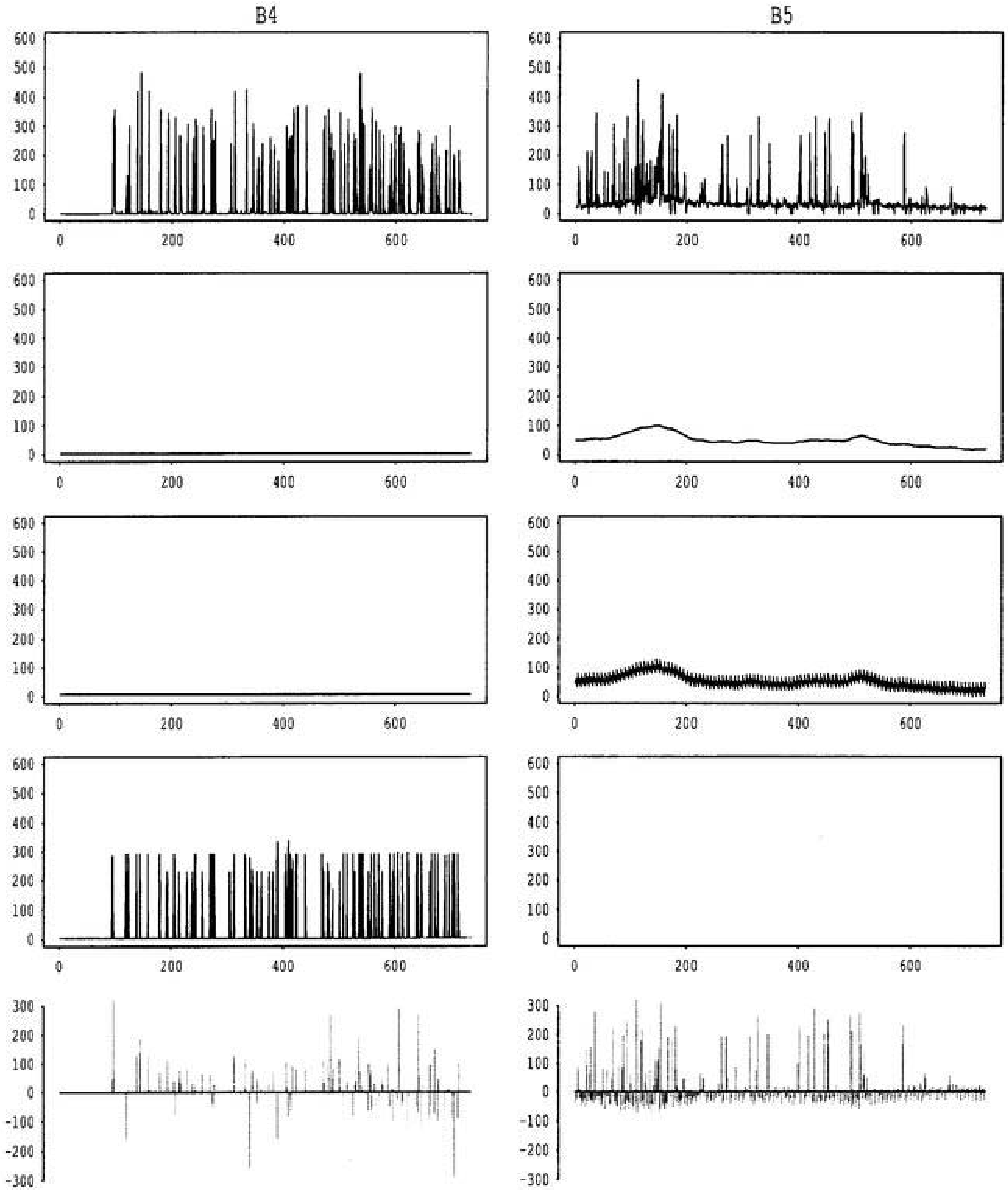


Figure 5.17: from top-to-bottom: the observation, fitted trend, trend + day-of-the-week, trend + day-of-the-week + time-varying explanatory variable component, residual component for store B

(1994/2/28 - 1996/3/3)

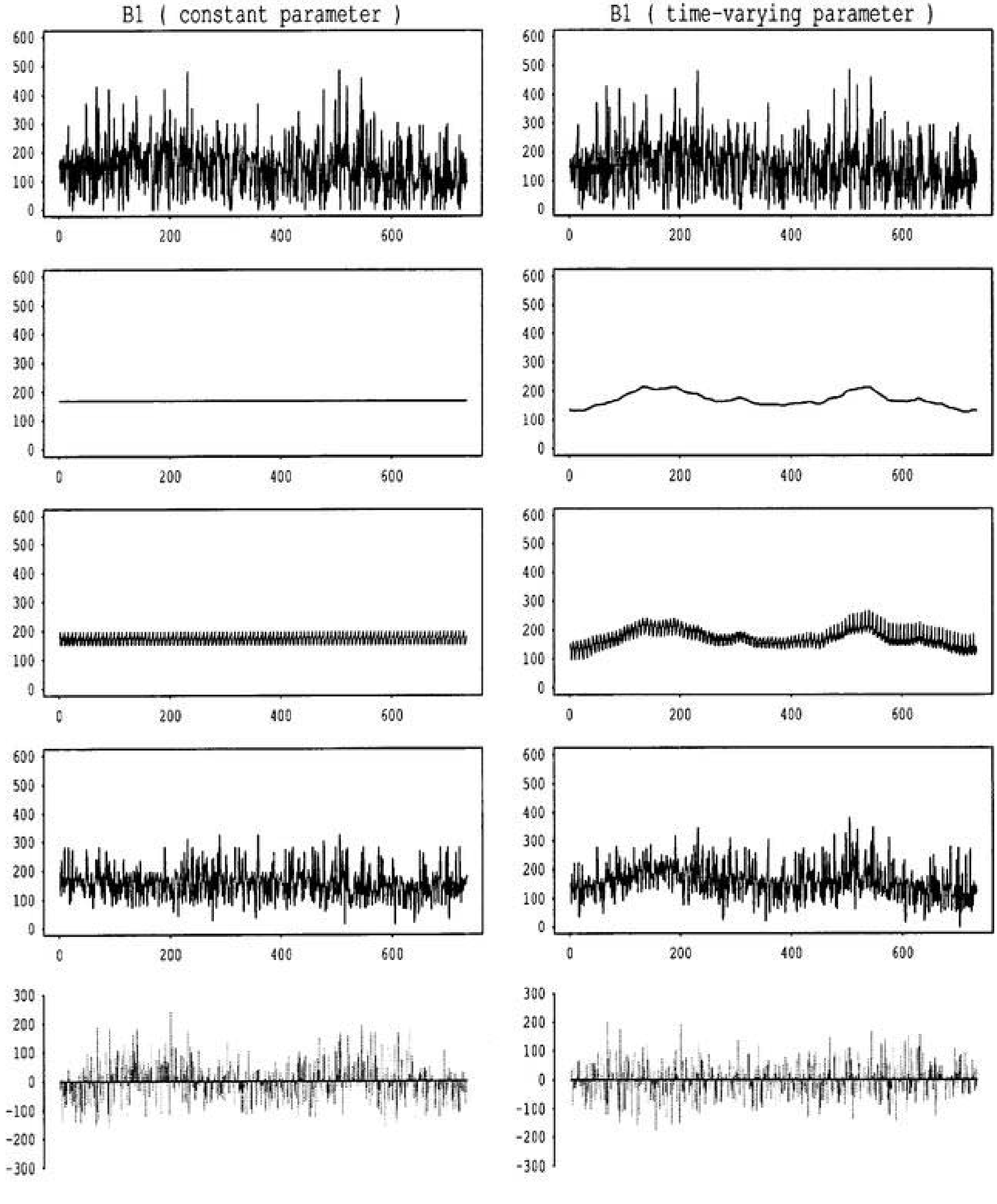


Figure 5.18: from top-to-bottom: the observation, fitted trend, trend + day-of-the-week, trend + day-of-the-week + constant (left) and time-varying (right) explanatory variable component, residual component for Brand B1 of store B
(1994/2/28 - 1996/3/3)

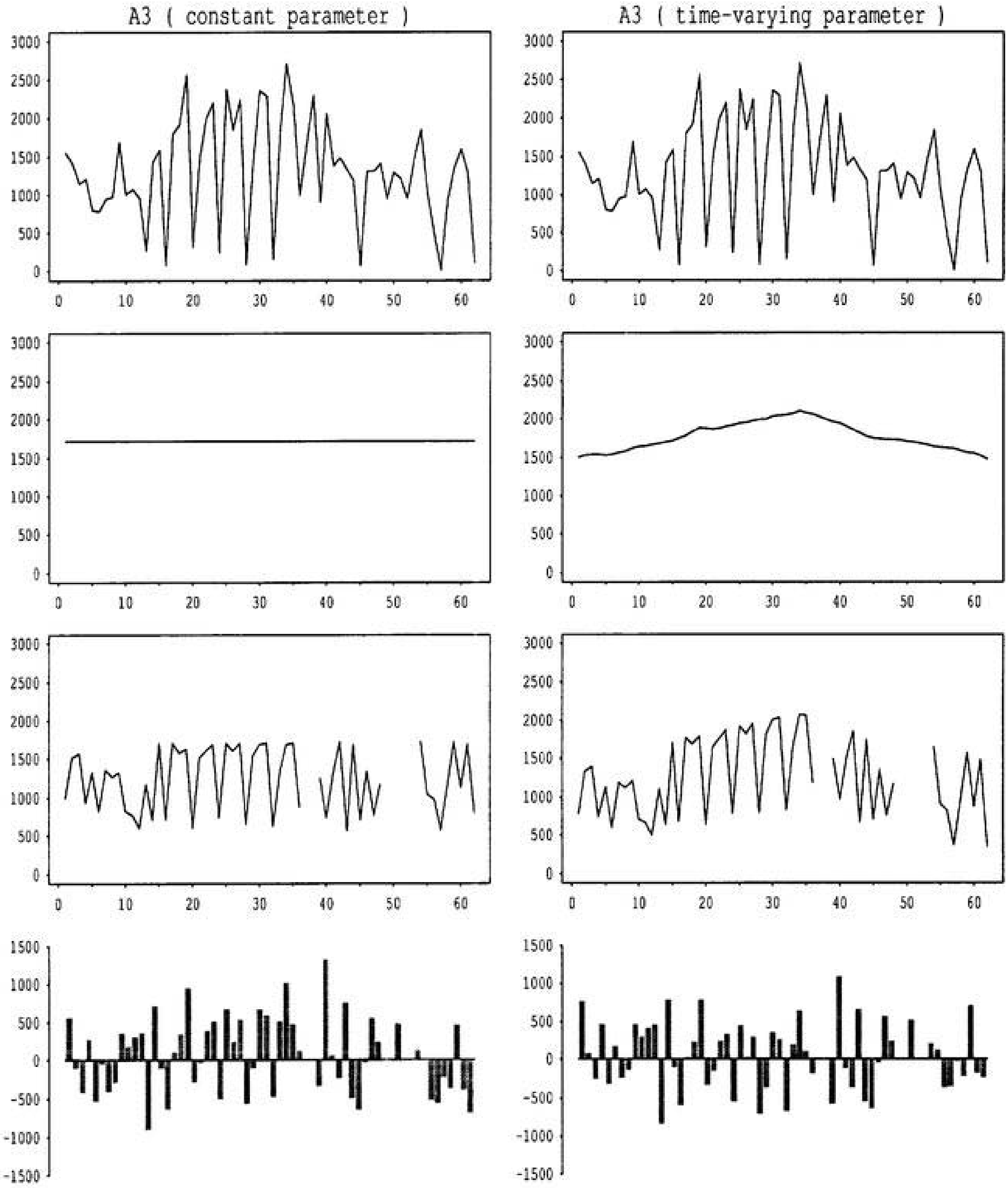


Figure 5.19: from top-to-bottom: the observation, fitted trend, trend + day-of-the-week, trend + day-of-the-week + constant (left) and time-varying (right) explanatory variable component, residual component for Brand A1 of store A
(1994/2/28 - 1996/3/3)

Chapter 6

6 Price Promotion Effect Decomposition into Brand Substitution Effect and Category Expansion Effect

Retail managers know from their past experiences that price promotion is often effective to cause high incremental sales of a promoted brand during the promotion period. Although they are interested in how much incremental sales are due to brand switching and how much incremental sales are due to category expansion (including store switching), there has been no means so far to decompose a price promotion effect into the one due to brand switching and that due to category expansion.

The purpose of this chapter is to demonstrate a method for decomposing an explanatory variable effect of a price promotion into the component due to brand switching and that due to category expansion, together with sales decomposition into long-term component of baseline sales (trend) and cyclical day-of-the-week effect. Our approach for this decomposition is to regress simultaneously brand sales and category sales on explanatory variables, by utilizing a Bayesian vector state space model, a unified model of time series analysis and regression analysis (Kondo and Kitagawa (1998)).

The idea to use brand sales data and category sales data is not new, as seen from cited articles (Neslin and Shoemaker(1983b); (Pindyck and Rubinfeld, 1981, Ch.11)) in Blattberg and Neslin (p.187, 1990). However, none of the two articles positively report on merits such as gains in estimation efficiency nor its usefulness of the scheme.

Our motivation to use brand sales data and category sales data comes from a necessity to decompose a price promotion effect into brand substitution effect and category expansion effect. In addition, this approach guarantees the results of category sales model are consistent with those of brand sales model, so that results are always comparable among brand sales and the category sales total. Therefore, our model can avoid a dilemma to decide which shoud believe between brand model and category model under a possible contradiction existing between them if separate model fittings on brand model and category model were conducted as questioned by Hanssens et al. (p. ,1990).

Multiple Exogeneous Effect Decomposition

The data used here are the same data from store A and store B in Chapter 5. The decomposition was conducted by using the best model obtained in Chapter 5.

The model 1 given in equation (6) was fitted to weekly data from store A. The model 2 that considers also day-of-the-week component given in equation (6) was applied to daily basis data from store B.

Model 1 (baseline + brand substitution + category expansion)

$$y(n) = t(n) + x(n) + w(n)$$

$$x(n) = g(n) + s(n)$$

Model 2 (baseline + day-of-the-week-effect + brand substitution + category expansion)

$$y(n) = t(n) + d(n) + x(n) + w(n)$$

$$x(n) = g(n) + s(n)$$

The values of the log-likelihood, AIC, and the number of parameters on the analysis of Model 1 were summarized in Table 6.24 where i denotes trend order and j denotes assumed period, X denotes explanatory variable component, G denotes brand substitution, and S denotes category expansion.

The results in Table 6.24 on store B data showed larger(?) log-likelihood on multiple exogeneous decomposition model. The attached tables exibits clear decomposition for each brand. The low fat type brand, B3, has rather independent movements and the incremental component contribute mostly to the category expansion. On other hand, for the private brand, B1, the incremental sales by the own price reduction did not take sales from competitors very much, but be taken by the competitors. The incremental sales could somewhat contribute to the category expansion. For national brand, B2, the incremental sales by the own price reduction could contribute to both the increments of the brand sales and the category expansion. The reduction in sales due to brand substitution effect by competitors was also recognized. The incremental sales by the own price reduction for brand b4 also could contribute to both the increments of the brand sales and the category expansion. This brand have not experience the reduction in sales due to brand substitution effect by competitors.

Table 6.24: Log-likelihood and AIC for model for store B data (T, W, S, G)

Model (T, W)	log-likelihood	AIC	No. of Parameters
Model (T, W)	-19512		60
Model (T, W, X)	-18259.8997		81
Model (T, W, S, G)	-18246.3383	††	91

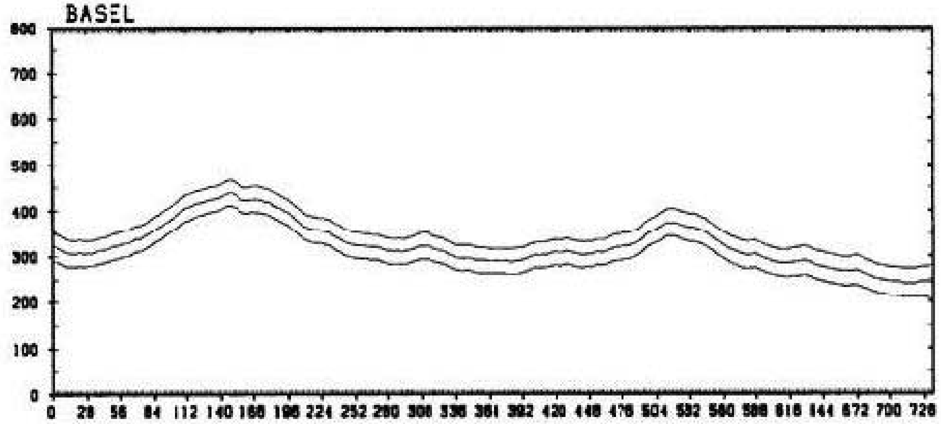
@REG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PROGRAM 2.1: @HIR13C.FORT OR 6

98-05-12 11 11 50 22

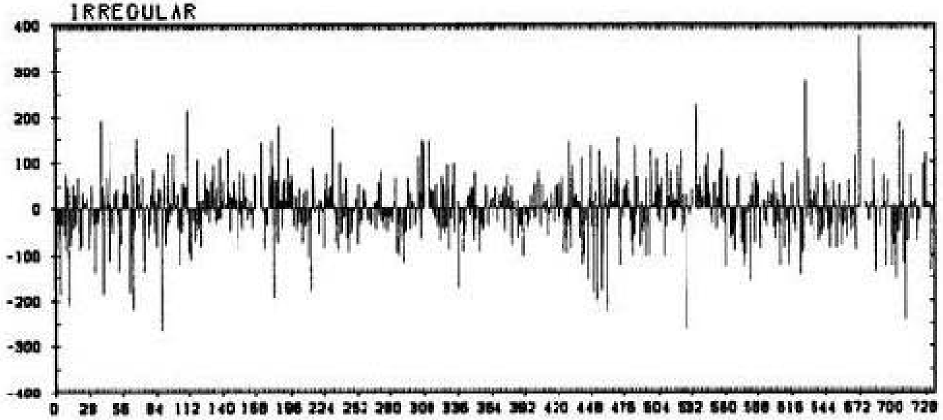
M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 0.000000	PERIOD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.049255	LOG-LH	= -18259.898438		

BASEL

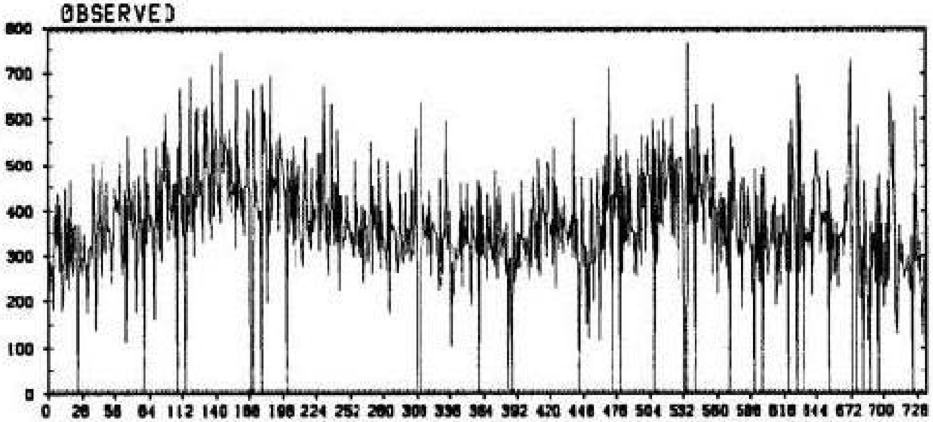


CYCLICAL

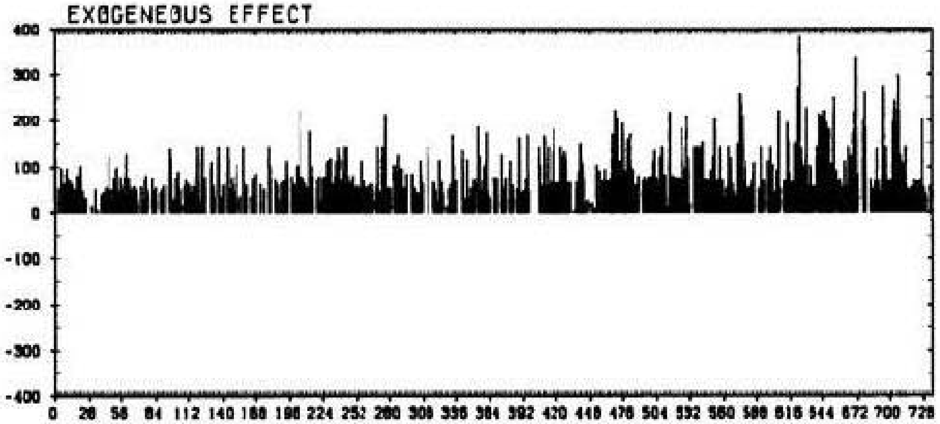
IRREGULAR



OBSERVED



EXOGENEOUS EFFECT

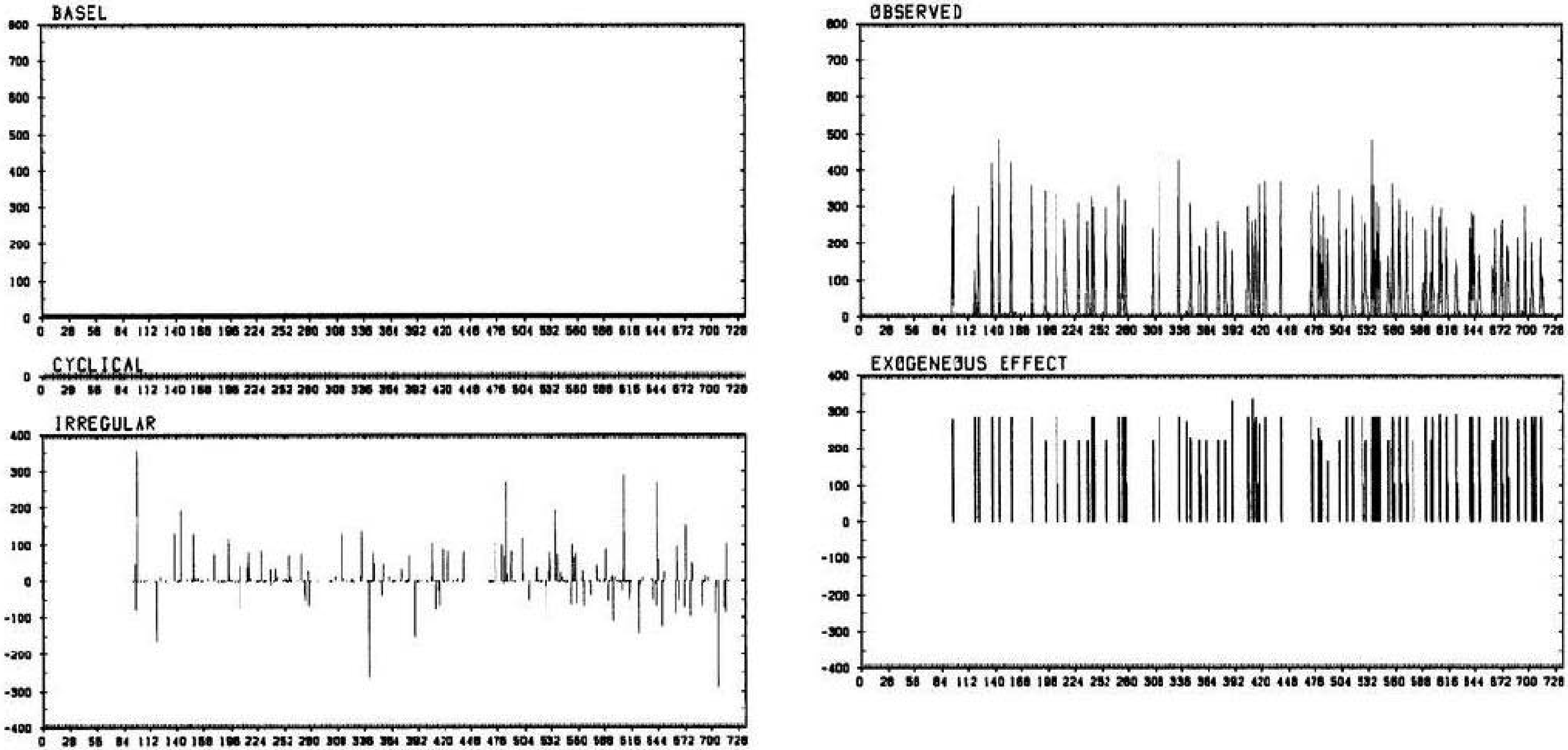


ØREG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PRØGRAM 2.1: GHIR13C.FØRT

98-05-12 11 11 58 26

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 0.000000	PERIØD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.049255	LØG-LH	= -18259.898438		



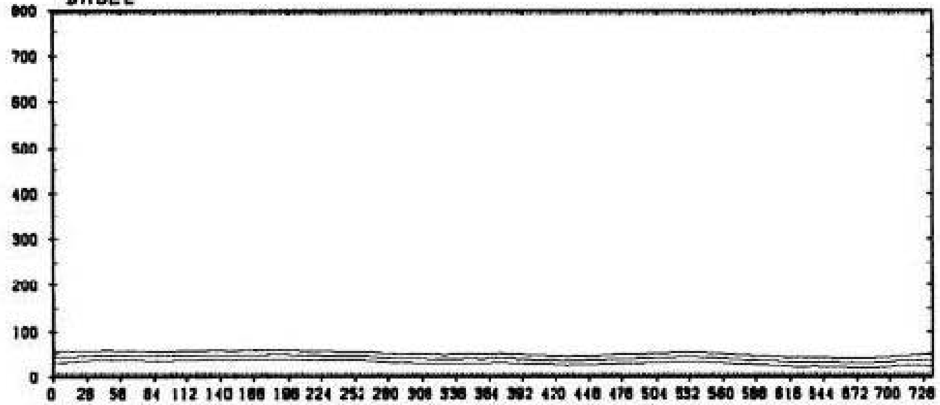
ØREG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PRØGRAM 2.1: ØHIR13C.FØRT

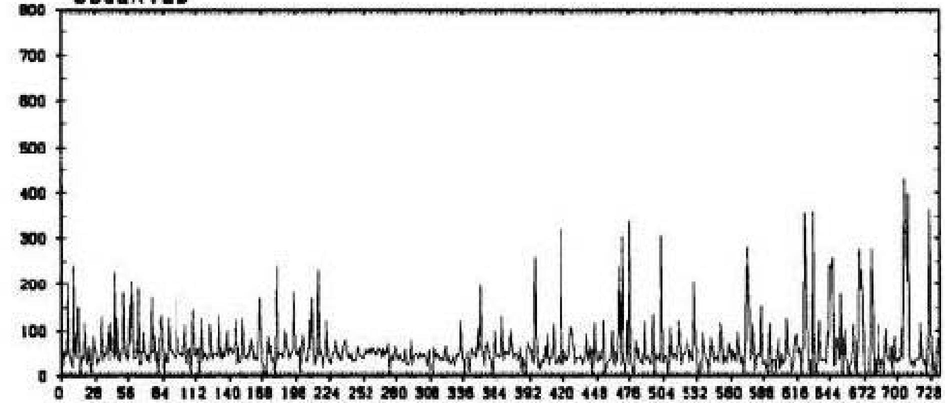
98-05-12 11 12 03 29

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 0.000000	PERIØD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.049255	LØG-LH	= -18259.898438		

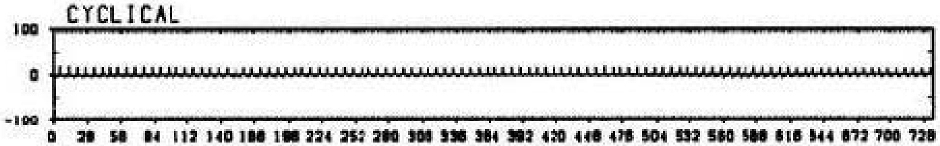
BASEL



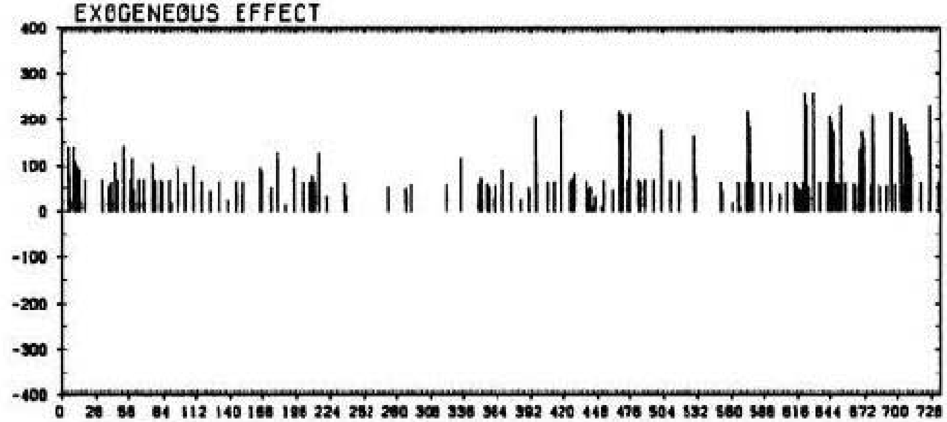
ØBSERVED



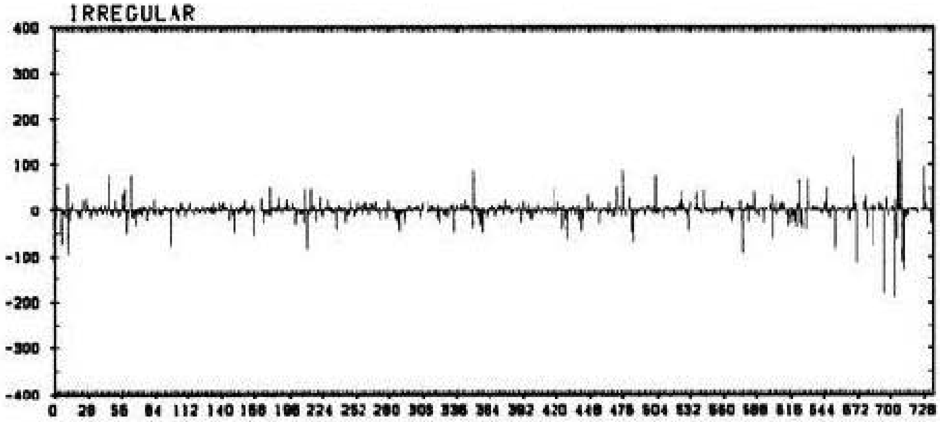
CYCLICAL



EXØGENEØUS EFFECT



IRREGULAR

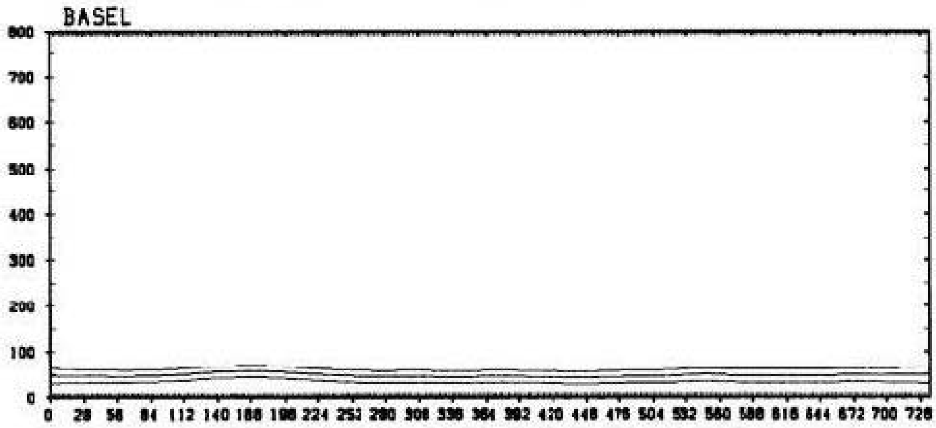


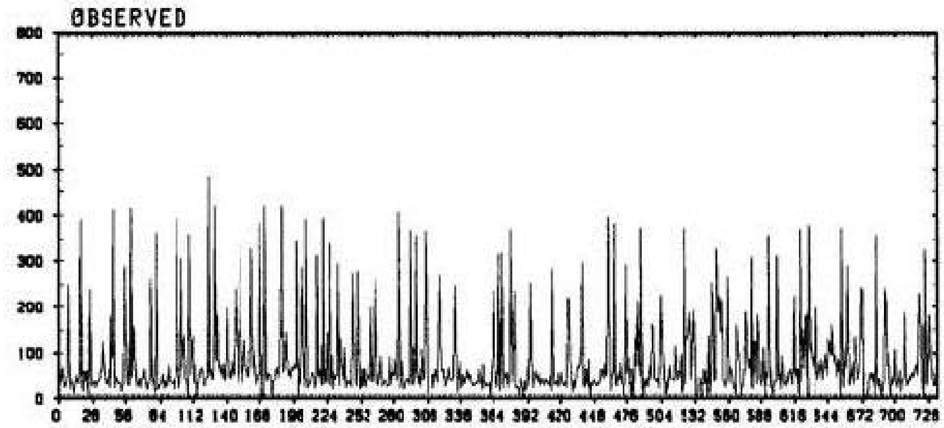
ØREG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

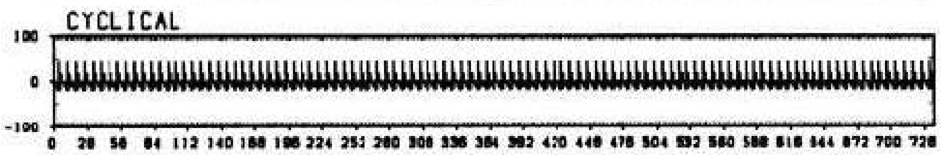
PRØGRAM 2.1: ØHIR13C.FØRT

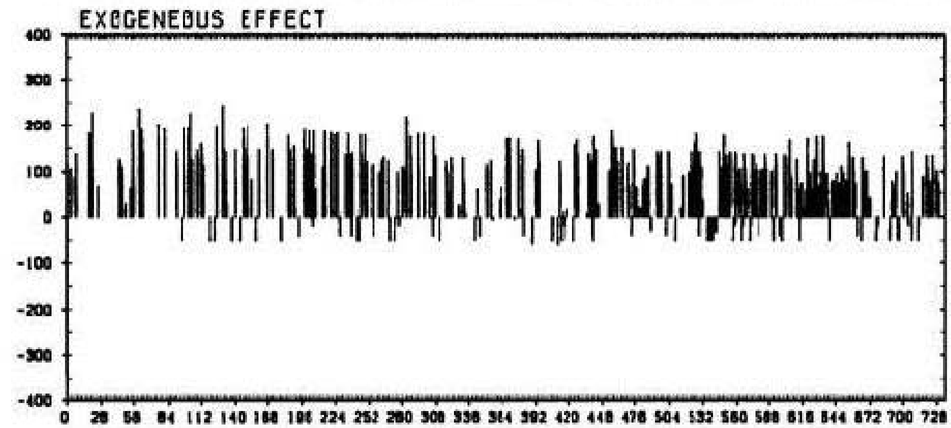
98-05-12 11 12 07 41

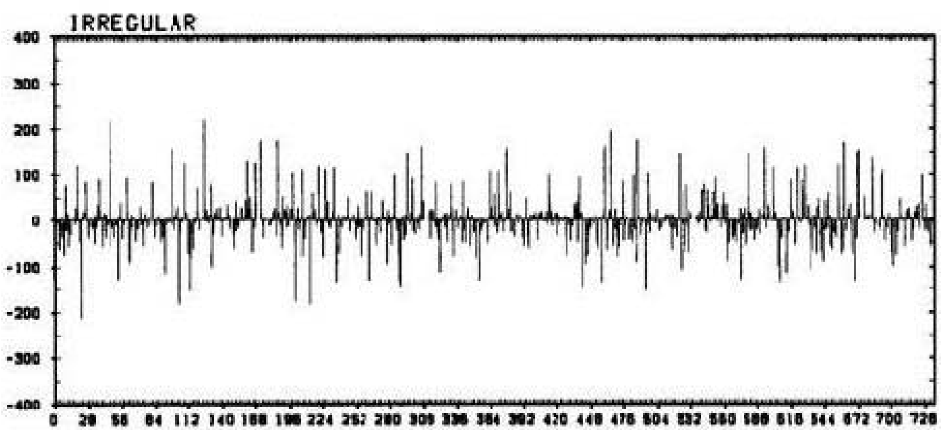
M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 0.000000	PERIØD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.049255	LØG-LH	= -18259.898438		









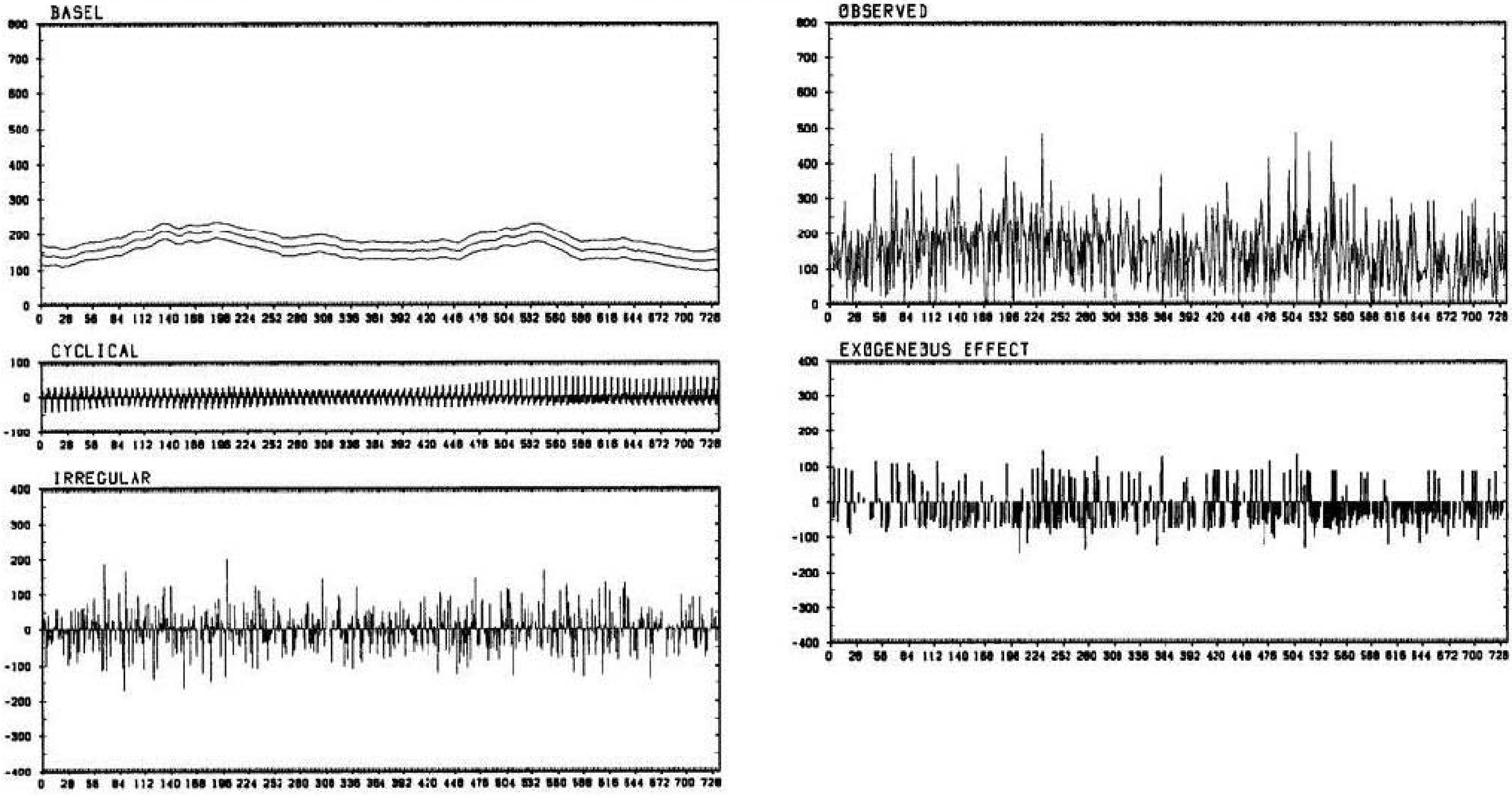


@REG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PROGRAM 2.1: @HIR13C.FORT

96-05-12 11 12 12 85

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 0.000000	PERIOD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.049255	LOG-LH	= -18259.898438		

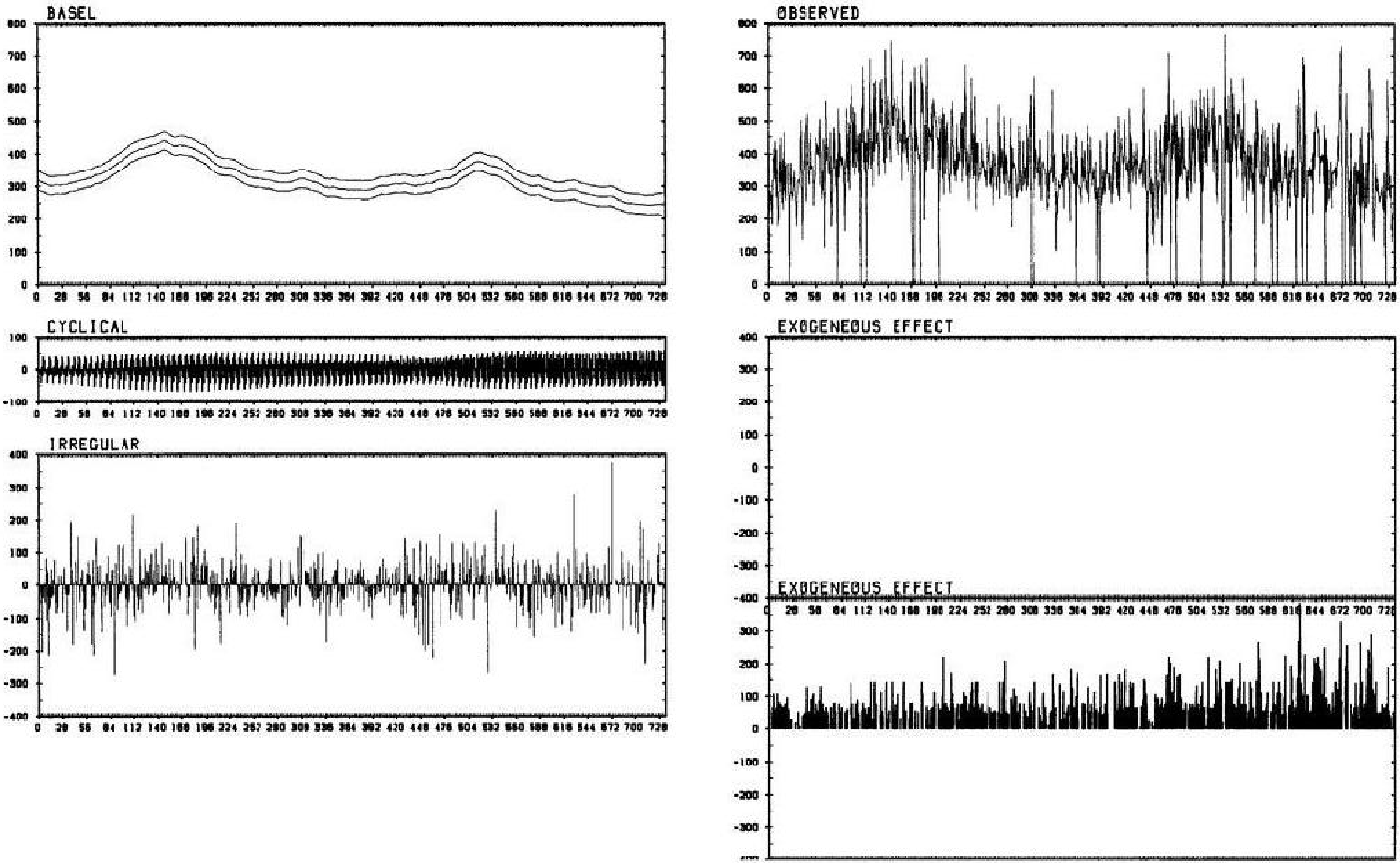


@REG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PROGRAM 2.1: @HIR13C.FORT PD

98-05-12 10 50 27 45

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 1.000000	PERIOD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.028161	LOG-LH	= -18246.335938		

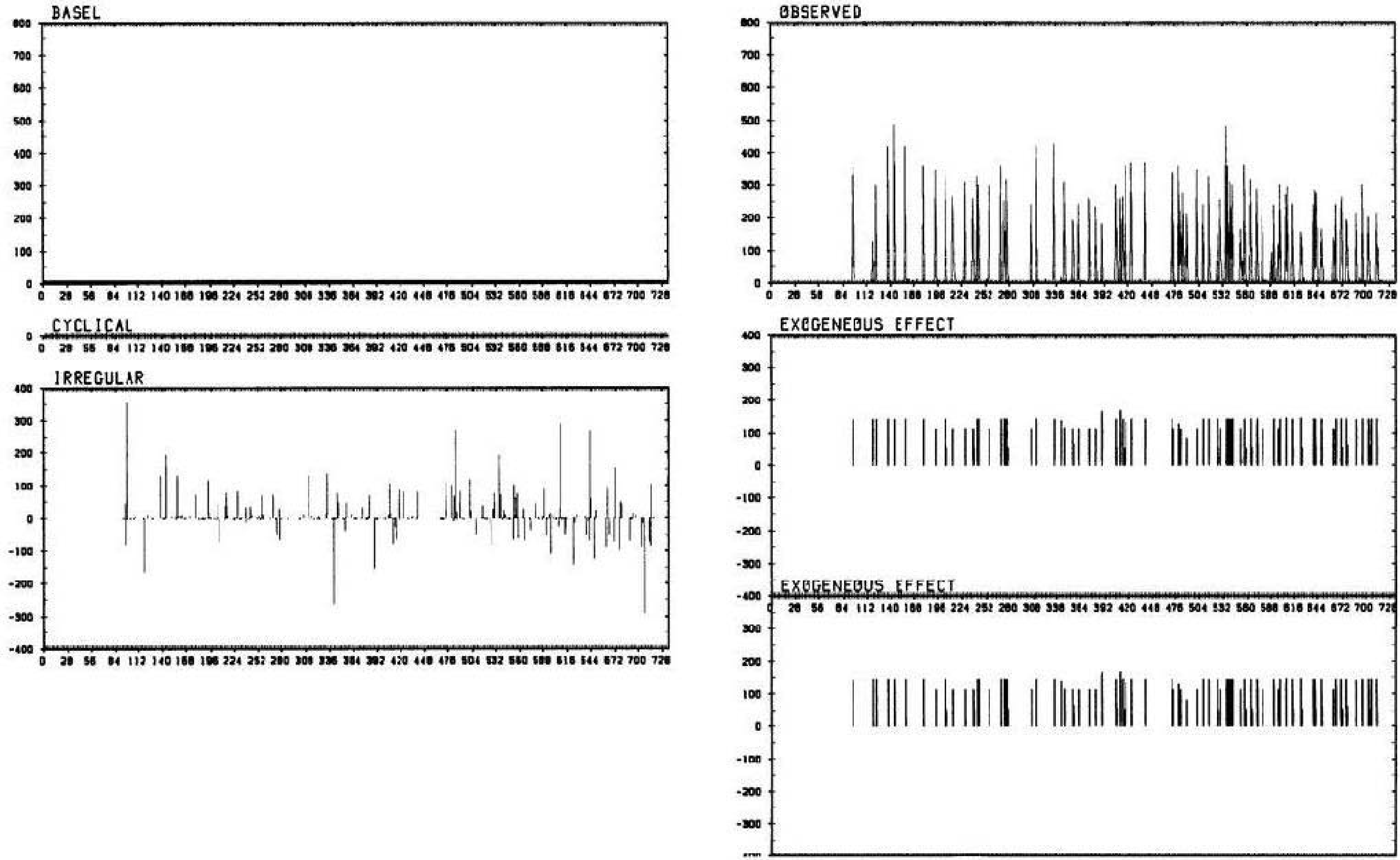


@REG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PROGRAM 2.1: @HIR13C.FORT

98-05-12 10 50 36 04

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 1.000000	PERIOD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.028161	LOG-LH	= -18246.335938		

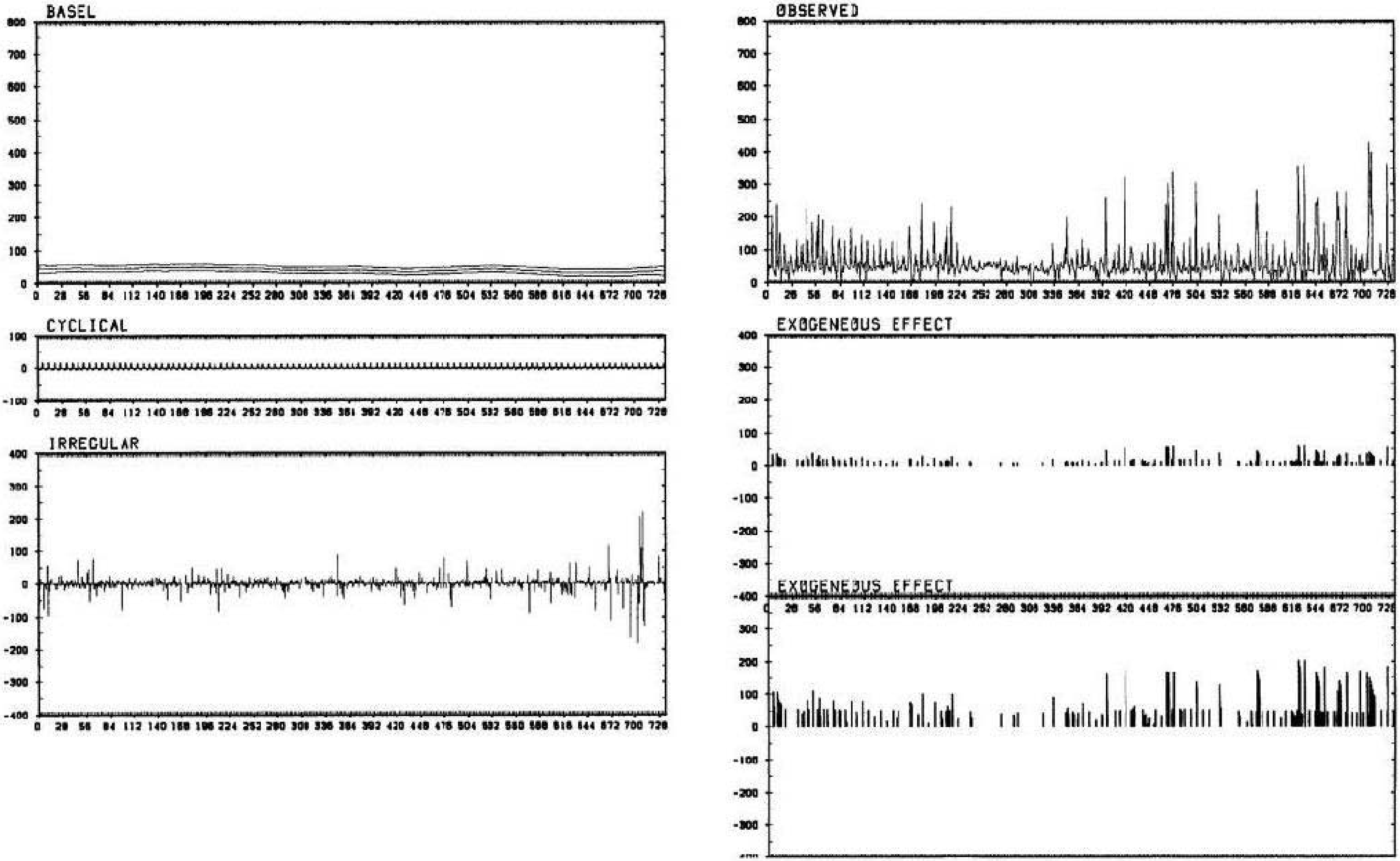


@REG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PROGRAM 2.1: @HIR13C.FORT

96-05-12 10 50 41 32

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 1.000000	PERIOD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.028161	LOG-LH	= -18246.335938		



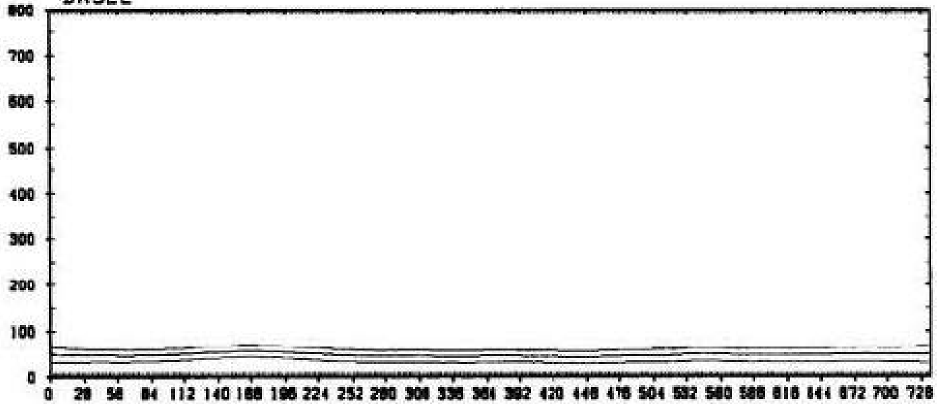
@REG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)

PROGRAM 2.1: @HIR13C.FORT

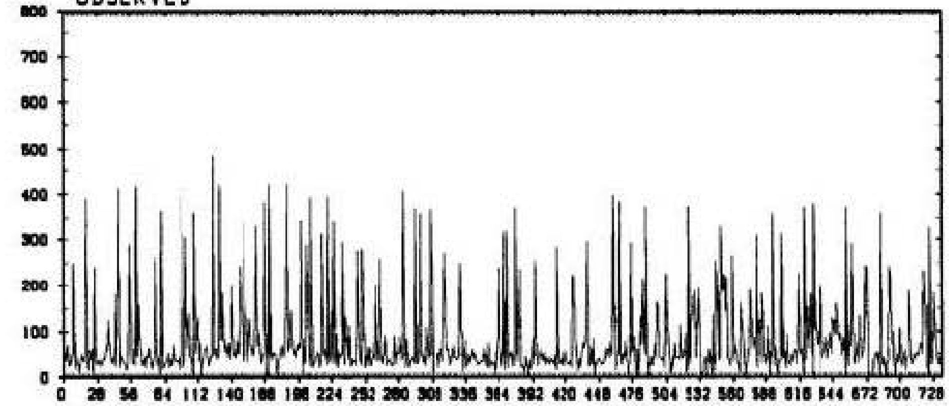
98-05-12 10 50 46 62

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 1.000000	PERIOD	= 7.000000	TAU1	= 3.883404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.028161	LOG-LH	= -18246.335938		

BASEL

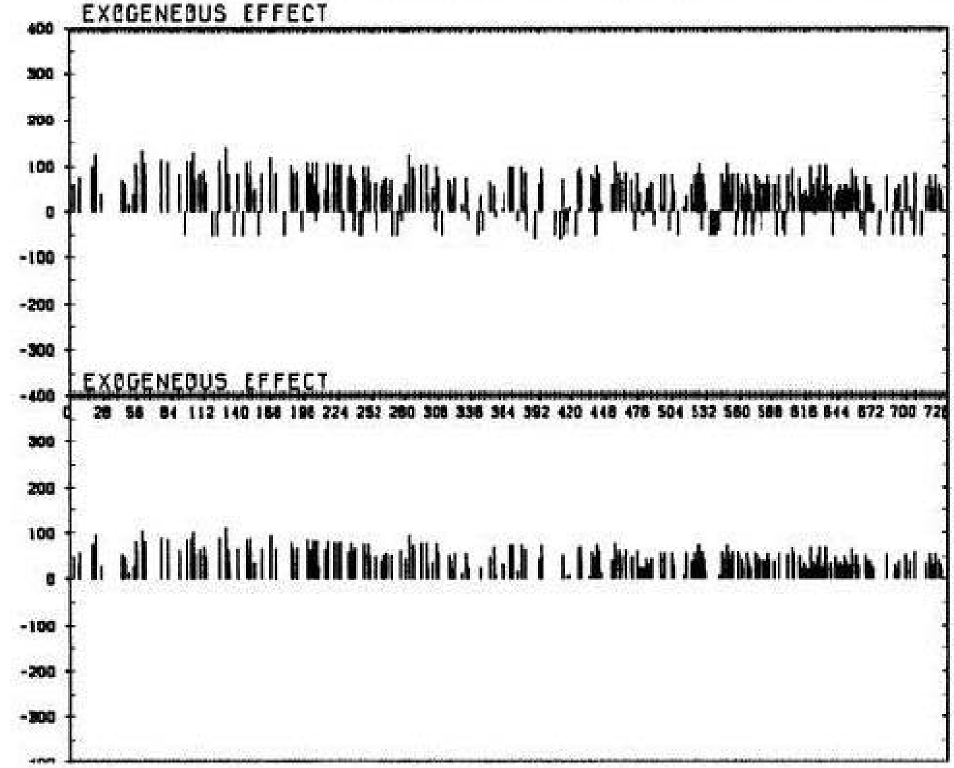


OBSERVED

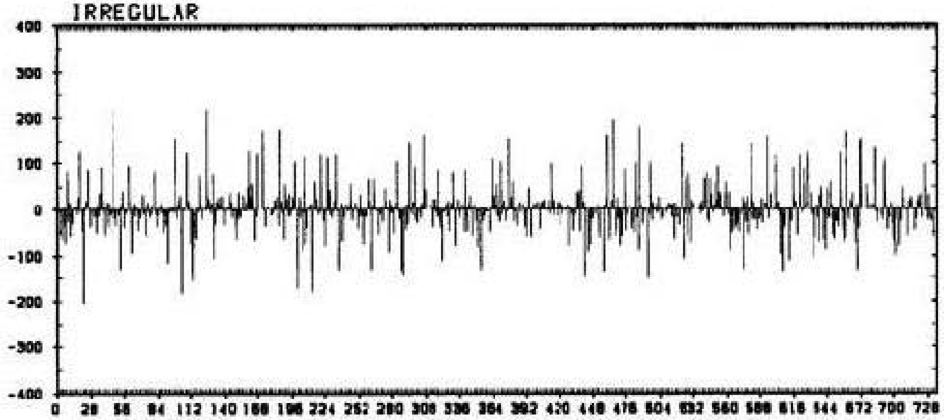


CYCLICAL

EXOGENEOUS EFFECT

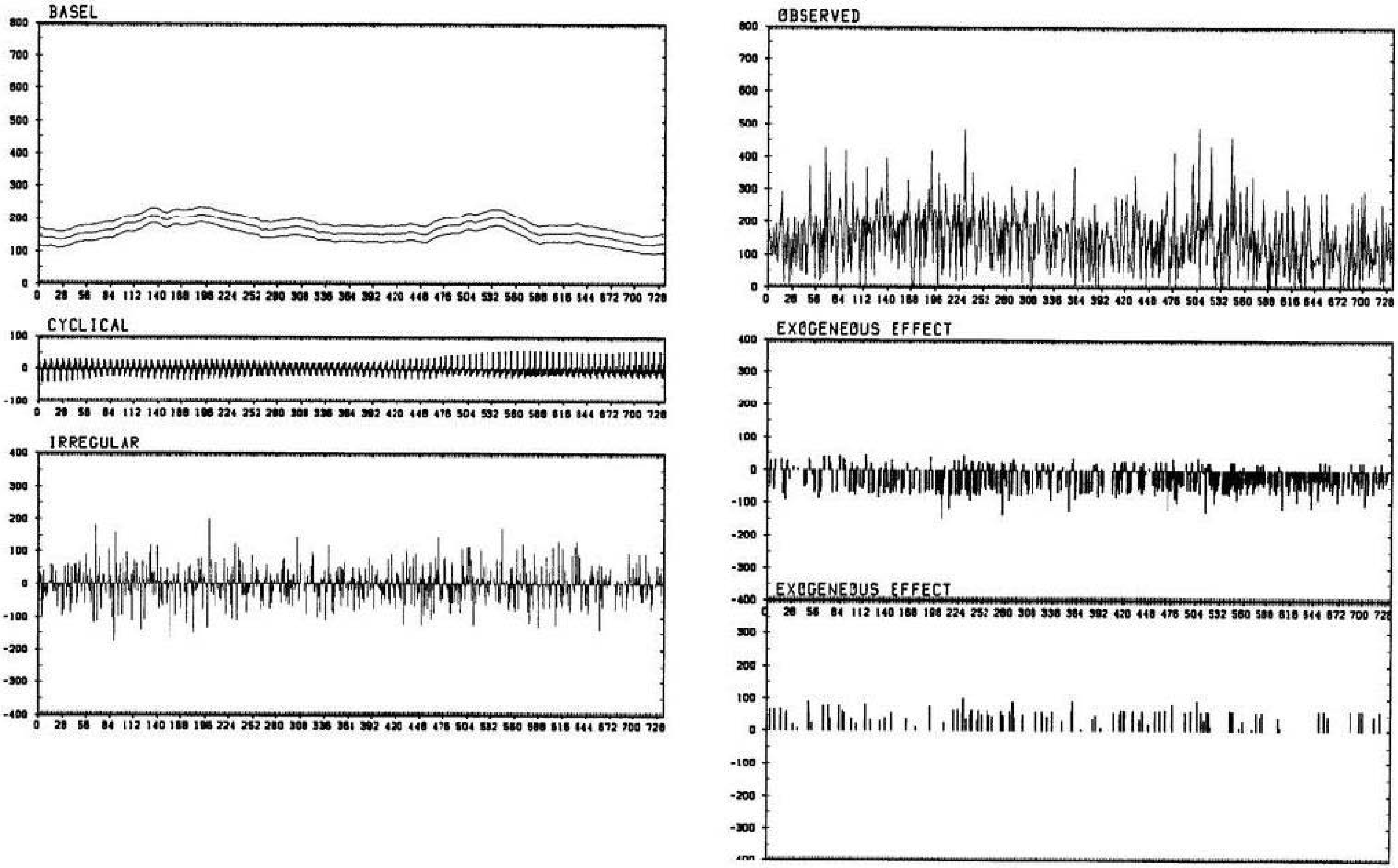


IRREGULAR



ØREG.DATA(P735) - B10 -128.0 -54.2, B4 -50.0 -30.0)
PRØGRAM 2.1: ØHIR13C.FØRT
98-05-12 10 50 51 17

M1	= 1.000000	M2	= 1.000000	M3	= 0.000000	M5	= 1.000000
M6	= 1.000000	PERIØD	= 7.000000	TAU1	= 3.880404	TAU2	= 1.152641
TAU3	= 1.152641	SIG2	= 5.028161	LØG-LH	= -18246.335938		



Chapter 7

7 Simultation Study

Simulaton was performed for the purpose of evaluationg trend, day-of-the-week effect, and explanatory variable effect with the following model:

$$y(n) = t(n) + d(n) + x(n) + w(n), \tag{7.75}$$

where $t(n)$, $d(n)$, $x(n)$ and $w(n)$ represent long-term baseline sales component, cyclical day-of-the-week effect component, short-term explanatory variable component, and observation noise, as specifed in equation 2.1. The number of brands are set to three and sample size as 300.

The observation noise $w(n)$ obeys the following normal distribution:

$$w(n) \sim N(\mathbf{0}\ ,\Sigma_w), \quad \Sigma_w = \begin{pmatrix} \sigma^2_{w_{11}} & & \mathbf{0} \\ & \sigma^2_{w_{22}} & \\ \mathbf{0} & & \sigma^2_{w_{33}} \end{pmatrix}. \tag{7.76}$$

Possible alternative observation noise is non-Gaussian.

Baseline component

$t(n) = (t_1(n), \ldots, t_3(n))^T$ is represented by the following $1st$ th order stochastic difference equation:

$$t(n) - t(n-1) = v_t(n), \tag{7.77}$$

where the system noise $v_{t_i}(n) = (v_1(n), \ldots, v_k(n))^T$ obeys the following Gaussian white noise

$$v_t(n) \sim N(\mathbf{0}\ ,\Sigma_t), \quad \Sigma_t = \begin{pmatrix} \sigma^2_{t_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma^2_{t_3} \end{pmatrix}, \tag{7.78}$$

Cyclical Day-of-the-week Component

$$\sum_{j=0}^{6} d(n-j) = v_d(n), \tag{7.79}$$

where the system noise $v_d(n) = (v_1(n), \ldots, v_3(n))^T$ obeys the following Gaussian white noise and corresponds to the change of the day-of-the-week pattern:

$$v_d(n) \sim N(\mathbf{0}\ , \Sigma_d), \quad \Sigma_d = \begin{pmatrix} \sigma_{d_1}^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_{d_3}^2 \end{pmatrix}. \tag{7.80}$$

Explanatory Variable component

$$g(n) = B(n)u^g(n), \tag{7.81}$$

where $u^g(n)$ is defined in the below as a price function for explanatory variable component and use actually the analyzed price fucntion. $B(n)$ represents time-varying coefficients of $u^g(n)$. The number of explanatory variables is set to 3

The source for time-variation is numerous as Cooley (1981) states. In our results in Chapter 5, we often saw the same (reversing) patterns between parameter in baseline component and incremental component in Figure 5.4? and Figure 5.5? . Therefore, we focus on the aspect that incremental component for a brand has usually a similar patern to basline component and sometimes a different pattern. We set up two scenarios as follows:

1) similar patern
2) similar patern + innovative patern

The time-varying parameter is expressed as

$$b_{ij}(n) = b_{ij}(0) + c_i t_i(n) + a_i(n), \tag{7.82}$$

where the value of constant competitive parameter was set at

$$B(0) = \begin{pmatrix} 0.6 & -0.5 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0 & 0.4 \end{pmatrix}, \tag{7.83}$$

and $0 < c_i < 1$ expresses a constant parameter as a ratio of baseline component and $a_i(n)$ is the following innovative pattern,

$$a_i(n) = f_i(n) \quad for \quad b_{11}(n), b_{12}(n), b_{22}(n) \tag{7.84}$$

$$a_i(n) = 0 \quad for \quad b_{33}(n). \tag{7.85}$$

Chapter 8

8 Summary and Conclusions

8.1 analysis 1

We propose a new methodology as a unified time series and econometric analysis with state space model to decompose category sales into component of local polynomial trend, the day-of-the-week effect, and explanatory variable effect. Like the Box-Jenkins ARIMA procedure, our approach allows a researcher to hypothesize possible models. However, a fundamental difference is in the process of model identification, parameter estimation, and diagnosis. Our approach uses AIC to select the best of alternative parametric models within and between model classes. The formula of AIC includes two essential components of *maximized log likelihood of the model* and *number of estimated parameters in the model* as bias correction. Therefore, necessary information to select the best model is already included in AIC. A researcher only has to set up possible alternative models and search for the model with the minimum value of AIC. In this sense, ours is essentially a semi-automatic extensive model alternative procedure.

On the other hand, Box-Jenkins approach requires extensive expert human intervention to achieve satisfactory modeling as follows: First, data is transformed to stationary time series data for example, by differencing, and integrating order is determined. After getting stationary data, the order of ARMA(p,q) model is determined to select several adequate models by using autocorrelation function (sometimes, extended autocorrelation function) and partial autocorrelation function. Thirdly, the parameters in the identified models are estimated by maximizing the (log) likelihood. Then, diagnostic checking and forecast follow.

Two example store level scanner data sets were analyzed to show the flexibility of our approach as a unified time series and econometric analysis. One involved only time series models with component of trend and cyclical day-of-the-week effect. The second included models of time series plus regression analysis on price deals. The both analysis gave the result that the inclusion of the day-of-the-week effect yielded a better model, accomplishing the purpose of a decomposition into the component of trend and day-of-the-week effect. On the second data set, the inclusion of explanatory component into time series model gave the best result. Here, several examples of linear or non-linear regression analyses were carried out.

8.2 analysis 2

This paper presents time-varying resgression coefficient estimates based on Bayesin vector state space model with Kalman filter. Kalman filter provides a simple and efficient algorithm to estimate and update time-varying parameters. The variances of the system noises in the stochastic process for random parameter is estimated by using Akaike's AIC criterion to determine the best parametric model among alternatives. Together with Mallow's Cp criterion, the AIC was demonstrated by Shibata (1980) to have an asymptotic minumum mean square one-step-ahead prediction error propertity, which are not the case in the consistent estimator schemes of Schwarz and Hannan and Quinn. The vector state space model has a great advantage for market response model. Since it is a unified model of time series analysis and regression analysis, there is no need for separate estimations of category total and brand share. The time series of each brand are always added up to that of the category total, which is a logically favourable point.

Three data sets of scanner sales were analyzed. The state estimates of time-varying regression coefficients were larger in all cases than the constant parameter cases. Together with our previous findings in Kondo and Kitagawa (1998), the effectiveness to derive incremental sales of its brand declines over time, the results poses serious doubt over the use of constant estimates for the analysis on longitudinal data of scanner sales. Further, results has shown that the extent of time-variation was large for private brands and the second large for national brands, both of which were affected by the price discount of competitive brands, making time-varying parameter models more appropriate than the constant ones. On the other hand, OLS estimation procedure would provide a reasonable estimates for brands which are not prone to seasonality or the influence of competitors price discounts.

Chapter 9

9 APPENDIX – For Future Research

9.1 Temporal Substitution Component Model

Temporal substitution component model describes a component which has large variaces for each period, but vanishes if all elements of the component are summed within a range of periods that temporal substitution effects remain. Therefore, it is characterised as "zero-sum effect" in the direction of period.

A temporal substitution component $s_\bullet(n) = \sum_{i=1}^{k} s_i(n)$ is defined as a category total component and is assumed to be expressed by an ARMA model,

$$s_\bullet(n) = \sum_{m=1}^{M} a_m s_\bullet(n-m) + \sum_{i=1}^{k}\sum_{j=0}^{q} c_i(n-j)u_i^s(n-j), \tag{9.86}$$

where $u^s(n) = (u_1^s(n), \ldots, u_k^s(n))^T$ is a price function for temporal substitution in the below, and a_m, $c_i(n)$, M, and q are the coefficient of autoregression, moving average, and the order, respectively.

The temporal substitution component is also expressed by

$$s_\bullet(n) = \sum_{j=0}^{\infty} h_j u^s(n-j), \tag{9.87}$$

where $h_j = (h_{j1}(n), \ldots, h_{jk}(n))$.

The structure of temporal substitution can be expressed by the restriction on the impulse response function,

$$\sum_{j=0}^{\infty} h_j = 0, \tag{9.88}$$

which is a nescessary constraint of the equation (9.86).

In practice, the effect of temporal substitution is assumed to remain not very long, so that temporal substitution may be simply expressed by a MA model. A stockpile effect is considered to be an effect which yields a large pulsive incremental sales followed by a negative response which cause a reduction in sales from the baseline sales. (Blattberg and Neslin (1990), pp.191). Specifically, for a stockpile effect with an *e-lagged* MA model, the equation (9.86) and (9.88) reduces to

$$s_{\bullet}(n) = \sum_{i=1}^{k} \sum_{j=0}^{q} c_i(n-j) u_i^s(n-j), \quad i = 1, \ldots, k, \tag{9.89}$$

$$\sum_{j=0}^{q} c_i(n-j) = 0, \quad c_i(n-j) = 0 \quad for \quad j = 1, \ldots, e, \tag{9.90}$$

where $e = q - 1$ is the number of lagged periods before negative responses appear due to stockpile effect. During the occurance of negative responses, the inputs of new price cuts are assumed to be not effective and, the sales, not responsive to the inputs. The lagged periods may be determined by multiplying the interpurchase time by the ratio of average extra purchses to baseline sales. If double quantity is purchased on average as a result of stockpile effect, the ratio is one, so that the length of the lagged periods equals to mean interpurchase time.

From the aforementioned past researches, temporal substitution is expected not to occur so often. There is a risk for a consumer to make a stockpile of rather unfavoured brands within a household, temporal substitution would not occur so often (Blattberg et al.(1981)). Only in the case that consumers think not to loose the oppurtunity, temporal substitution occurs. In addition, if a consumer who usually purchases a category at a different store in the next purchasing occasion, his or her purchasing a double quantity is accounted for a store substitution. From this point of view, a price cut to cause temporal substitution must be substantial, for example, a half of the regular price.

The price function of temporal substitution effect, $u^s(n) = (u_1^s(n), \ldots, u_k^s(n))^T$, is defined as follows:

$$
\begin{aligned}
u^s(n) &= \tilde{u}(n) \quad if \quad \tilde{u}(n) + Th^s \leq 0 \\
&= 0 \quad \text{otherwise},
\end{aligned} \tag{9.91}
$$

where $Th^s \geq 0$ determines a threshold level to cause temporal substitution.

If an ARMA model is used to achieve a parsimony of parameters, how to place a necessary constraint for temporal substitution model is an area to need further research.

9.2 Constraint Condition for Stockpile Effect Component

$$
\Gamma \begin{pmatrix} s_1(n) \\ s_2(n) \end{pmatrix} = \Gamma \begin{pmatrix} c_1(n) & 0 \\ 0 & c_2(n) \end{pmatrix} \begin{pmatrix} u_1^s(n) \\ u_2^s(n) \end{pmatrix} = \begin{pmatrix} c_1(n) & 0 \\ c_1(n) & c_2(n) \end{pmatrix} \begin{pmatrix} u_1^s(n) \\ u_2^s(n) \end{pmatrix}, \tag{9.92}
$$

$$
c_i(n-1-e)u_i^s(n-1-e) + c_i(n)u_i^s(n-1-e) = 0, \quad i = 1,2, \tag{9.93}
$$

where $e = q - 1$ is the number of lagged periods to appear a negative lagged responses after a large pulsive incremental sales.

$$
\Gamma \begin{pmatrix} s_1(n) \\ s_2(n) \end{pmatrix} = \begin{pmatrix} c_1(n) & c_1(n-1-e) & 0 & 0 \\ c_1(n) & c_1(n-1-e) & c_2(n) & c_2(n-1-e) \end{pmatrix} \begin{pmatrix} u_1^s(n) \\ u_1^s(n-1-e) \\ u_2^s(n) \\ u_2^s(n-1-e) \end{pmatrix} \tag{9.94}
$$

9.3 System Model for Stockpile Effect with $e - lagged$ MA model for $e = 2$

The system model can be written as follows:

$$
c(n) = \begin{pmatrix} c_1(n) \\ c_1(n-1) \\ c_1(n-2) \\ c_1(n-3) \\ c_2(n) \\ c_2(n-1) \\ c_2(n-2) \\ c_2(n-3) \end{pmatrix} = \begin{pmatrix} 1 &&&&&&& \\ & 1 &&&&&& \\ && 1 &&&&& \\ &&& 1 &&&& \\ &&&& 1 &&& \\ &&&&& 1 && \\ &&&&&& 1 & \\ &&&&&&& 1 \end{pmatrix} \begin{pmatrix} c_1(n-1) \\ c_1(n-2) \\ c_1(n-3) \\ c_1(n-4) \\ c_2(n-1) \\ c_2(n-2) \\ c_2(n-3) \\ c_2(n-4) \end{pmatrix}, \tag{9.95}
$$

$$\Gamma H_c = \begin{pmatrix} u_1^s(n) & 0 & 0 & u_1^s(n-3) & & & & \\ u_1^s(n) & 0 & 0 & u_1^s(n-3) & u_2^s(n) & 0 & 0 & u_2^s(n-3) \end{pmatrix}, \tag{9.96}$$

$$\Gamma H_s c(n) = \begin{pmatrix} c_1(n)u_1^s(n) + c_1(n-3)u_1^s(n-3) \\ c_1(n)u_1^s(n) + c_1(n-3)u_1^s(n-3) + c_2(n)u_2^s(n) + c_2(n-3)u_2^s(n-3) \end{pmatrix}, \tag{9.97}$$

with

$$c_i(n-3)u_i^s(n-3) + c_i(n)u_i^s(n-3) = 0, \quad i = 1, 2. \tag{9.98}$$

References

[1] Adams, A.J. and M.M. Moriarty (1981), "The Advertising Sales Relationship: Insights from Transfer Function Modeling, *Journal of Advertising Research*, 21(June), 41–46.

[2] Akaike, H. (1980a), "Likelihood and the Bayes procedure", in *Bayesian Statistics*, J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith, A.F.M. eds., University Press, Valencia, 143–166.

[3] Akaike, H. (1980b), "Seasonal Adjustment by a Bayesian Modeling", *J. Time Series Anal.*, 1,1–13.

[4] Anderson, B.D.O. and J.B. Moore (1979), *Optimal Filtering*, Prentice Hall, New Jersey.

[5] Athans, M. (1974), "The Importance of Kalman Filtering Methods for Economic Systems", *Annals of Economic and Social Measurement*, 3(1), 49–64.

[6] Blattberg, R.C., G.D. Eppen, and J. Lieberman (1981), "A Theoretical and Empirical Evaluation of Price Deals for Consumer Nondurables", *Journal of Marketing*, 45, no.1, 116–129.

[7] Blattberg, R.C. and S.A. Neslin (1990), *Sales Promotion: Concepts, Methods, and Strategies*, Prentice Hall, New Jersey.

[8] Box, G.E.P. and G.C. Tiao (1973), *Basian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA.

[9] Box, G.E.P. and G.C. Tiao (1975), "Intervention Analysis with Application to Economic and Environmental Problems," *Journal of the American Statistical Association*, 70, 70–79.

[10] Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, second edition. San Francisco: Holden-Day.

[11] Bozdogan, H. (1994), *Proceedings of the First US/Japan Conference on the Frontier of Statistical Modeling: An Informational Approach*, Kluwer Academic Publishers.

[12] Doyle, P. and J. Saunders (1985), "The Lead Effect in Marketing," *Journal of Marketing Research*, 22, no.1, 54–65.

[13] Duncan, D.B. and S.D. Horn (1972), "Linear Dynamic Recursive Estimation from Viewpoint of Regression Analysis", *J. Amer. Statist. Assoc.*, 67, 815–821.

[14] Duncan, G., W.L. Gorr, and J. Szczypula (1993), "Bayesian Forecasting for Seemingly Unrelated Time Series: Application to Local Government Revenue Forecasting", *Management Science*, 39, 275–293.

[15] Gersch, W. and G. Kitagawa (1983a), "The Prediction of Time Series with Trends and Seasonalities", *Journal of Bus. Econ. Statist*, 1, 253–264.

[16] Hanssens, D.M., L.J. Parsons, and R.L. Schultz (1990), *Market Response Models: Econometric and Time Series Analysis*. Boston: Kluwer Academic Publishers.

[17] Harrison, P.J. and C.F. Stevens (1976), "Bayesian forecasting", (with discussion), *J. R. Statist. Soc.*, B38, 205–247.

[18] Harvey, A.C. (1989), *Forecasting. Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Victoria, Australia.

[19] Harvey, A.C. and G.D.A. Phillips (1980), "The Estimation of Regression Models with Time-varying Parameters", *Proceedings of a Sumposium in Jonour of Oskar Morgenstern*, Vienna.

[20] Helmer, R.M., and J.K. Johansson (1977), "An Exposition of the Box-Jenkins Transfer Function Analysis with an Application to the Advertising-Sales Relationship", *Journal of Marketing Research*, 14, no.2, 227–239.

[21] Jeffreys, H. (1961), *Theory of Probability, third edition*, Oxford University Press, Oxford.

[22] Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl, and Tsoung-Chao Lee (1985), *The Theory and Practice of Econometrics*, 2nd edition, John-Willey & Sons, New York.

[23] Kalman, R.E. (1960), "A new Approach to Linear Filtering and Prediction Problems ", *Trans. Amer. Soc. Mech. Eng., J. Basic Engineering*, 82, 35-45.

[24] Kalman, R.E. and R.S. Bucy (1961), "New Results in Linear Filtering and Prediction Theory", *Trans. of the ASME, J. Basic Engng.*, 83D, 95-108.

[25] Kitagawa, G. (1981), "A Nonstationary Time Series Model and its Fitting by a Recursive Filter", *J. Time Series Anal.*, 2, 103–116.

[26] Kitagawa, G. and W. Gersch (1984), "A Smoothness Priors - State Space Modeling of Time Series with Trend and Seasonality", *Journal of the American Statistical Association*, 79, No. 386, 378–389.

[27] Kitagawa, G. and W. Gersch (1985b), "A smoothness Priors Time-varying AR Coefficient Modeling of Nonstationary Covariance Time Series", *IEEE Trans. Automat. Contr.*, AC-30, No.1, 48–56.

[28] Kitagawa, G. (1993), *Fortran 77 Programming for Time Series Analysis*, (in Japanese), (Iwanami, Tokyo).

[29] Kitagawa, G. and W. Gersch (1996), *Lecture Notes in Statistics 116: Smoothness Priors Analysis of Time Series*, Springer-Verlag, New York.

[30] Leone, R.P. (1983), "Modeling Sales-Advertising Relationship: An Integrated Time Series Econometric Approach," *Journal of Marketing Research*, 20, 291–295.

[31] Leone, R.P. (1987), "Forecasting the Effect of an Environmental Change on Market Performance: An Intervention Time-Series Approach", *International Journal of Forecasting*, 3, 463–478.

[32] Lewis, L.F. (1986), *Optimal Estimation*. New York: John Wiley & Sons.

[33] Liu, Lon-Mu, and D.M. Hanssens (1981), "A Bayesian Approach to Time-varying Cross-sectional Regression Models", *Journal of Econometrics*, 15, 341–356.

[34] Lindley, D.V. and A.F.M. Smith (1972). Bayes Estimates for the Linear Model (with discussion), *J. Roy. Statist. Soc. Ser. B*, 34, 1–41.

[35] Maritz, J.S. and T. Lwin (1989). *Empirical Bayes Methods, Second Edition*, Chapman and Hall, London.

[36] Montgomery, D. and L. Johnson (1976). *Forecasting and Time Series Analysis*. New York: McGraw Hill.

[37] Moriarty, M. M. (1985), "Retail Promotional Effects on Intra and Interbrand Sales Performance", Journal of Retailing, 61, 3, 27–47.

[38] Morrison, G.W. and D.H. Pike (1977), "Kalman Filtering Applied to Statistical Forecasting", *Management Science*, 23, 768–778.

[39] Schweppe, C.F. (1965), "Evaluation of Likelihood Functions for Gaussian Signals", *IEEE Transactions on Information Theory*", IT -11, 61-70.

[40] Shiller, R. (1973). "A Distributed Lag Estimator Derived from Smoothness Priors", *Econometrica*, 41, 775–788.

[41] Slade, M.E. (1989), "Modeling Stochastic and Cyclical Components of Technical Change", *Journal of Econometrics*, 41, 363–383.

[42] Stone, M. (1977), "Comments on Model Selection Criteria of Akaike and Schwartz," *Journal of the Royal Statistical Society Series B*, 41(2), 276–278.

[43] Tegene, A. (1990), "The Kalman Filter Approach for Testing Structural Change in the Demand for Alcoholic Beverages in US", *Applied Economics*, 22, 1407–1416.

[44] Tegene, A. (1991), "Kalman Filter and the Demand for Cigarettes", *Applied Economics*, 23, 1175–1182.

[45] West, M., P.J. Harrison, and H.S. Mingon (1985). Dynamic Generalized Linear Models and Bayesian Forecasting (with discussion), *J.Amer. Statist. Assoc.*, 80, 73–97.

[46] Whittaker, E.T. (1923),"On a New Method of Graduation", *Proc. Edinborough Math. Assoc.*, 78, 81–89.

[47] Wichern, D.W. and R.H. Jones (1977), "Assessing the Impact of Market Disturbances Using Intervention Analysis, *Management Science*, 24, 329–337.

[48] Xie, J., X. M. Song, M. Sirbu, and Q. Wang (1997)," "Kalman Filter Estimation of New Product Diffusion Models" *Journal of Marketing Research*, 34, 378–393.