東　　　海　　　大　　　學

工業工程與經營資訊研究所

博士論文

以知識引導之調適型基因演算法預測模式--以消費性包裝產品業為例

博　士　生：王金鍊

指導教授：王立志 教授

中 華 民 國 一 OO 年 一 月

A Knowledge Guided Adaptive GA based Forecasting Model in Consumer Packaged Goods Industry

By
Chin-Lien Wang

Advisor: Prof. Li-Chih Wang

A Thesis
Submitted to the Institute of Industrial Engineering and Enterprise Information at
Tunghai University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in
Industrial Engineering and Enterprise Information

January 2011
Taichung , Taiwan , Republic of China

以知識引導之調適型基因演算法預測模式—以消費性包裝產品業為例

學生：王金鍊　　　　　　　　　　　　　指導教授 :王立志 博士

東海大學工業工程與經營資訊研究所

摘 要

由於激烈競爭，消費性包裝產品(CPG)生產業者多利用降價及各種非價格促銷活動來維繫競爭力。然而由促銷所造成的銷售量劇烈變動，以及特定促銷方式與農曆節日等稀疏的觀察資料，常讓統計型方法捉襟見肘，即使統計型方法一般的預測績效相當可靠，它們大多無法利用背景資訊進行預測。相對的，判斷型預測由於可以利用背景資訊，因而在業界廣為流行，卻有諸多偏誤與不一致的問題。分解式迴歸模型似乎成為整合上述兩種方法的自然選擇，以其可以納入背景因子如促銷與節日效應於模型，以充份利用背景資訊，並以「化整為零，各個擊破」的方式有效舒解預測人員的資訊超載負擔。可是這些背景因子常是特殊事件，它們的歷史資料稀少資訊量不足，導致共線性問題的滋生，重要變數被移除，並讓最小平方的參數估計失真。

在本論文的第一階段，使用了由知識引導的調適型基因演算法（KGAGA）來估計參數，在目標函數使用 MAPE 而非一般常用的 MSE 來評估參數的可能解，以減輕逸出值的平方計算所造成的不良影響，並依據專業知識設定參數限制範圍，使所求得的參數更有意義並且更加真實。 值得一提的是，本論文提出一種有效監測與跳脫區域解陷阱的機制（DEMA)，以最近 l 期最佳解目標函數值改善的移動平均數（$MAFI(l)$），偵測區域陷阱，並以廣域搜尋突變運算子與近域搜尋突變運算子組成的迴圈，依特定機率在個體解染色體多個位置執行位元對調，以快速增進族群的差異性，跳脫區域陷阱，同時藉由近域搜尋突變運算子的極低突變率與高雜交率以維繫搜尋的收斂能力，搭配正常搜尋達成再次收斂。父代與母代隨機選自目標函數值不同的族群，可紓解部份”選擇最適”所衍生的族群差異快速降低的困擾。KGAGA 的求解速度快於普通 GA，所求得個體解的品質也穩定優於普通 GA.

倘若預測期有任何可預期的變動未能由模型有效處理，本論文的第二階段提出一個機械式調整機制，由一組額外的方程式，包括年節前後季節指數的重新定位、比例調整及綜合調整等三種調整方式負責處理。此外也將週末效應列入考量， 穩定而有效精算出相關背景因子在預期的時間偏移後，節日效應的重新定位以及預期變動後促銷與假日等的混合效果，不涉及人為主觀判斷，以消除判斷式調整常見的偏誤與不一致現象。

針對國內一家 CPG 品牌業者通路商的實證研究顯示，本論文使用的 KGAGA 在面對背景因子時會是較 OLS 及 OGA 更好的模型擬合選擇，不管在參數估計及多期樣本外預測都有更好表現。KGAGA 在樣本外多期預測的表現也優於 ARIMA，指數平滑，及天真法。此外也證實本論文使用的機械式調整機制能有效降低各種淡旺季週預測之 MAPE，而主要的改善來自大型調整。

關鍵字詞：背景資訊；預測模式；基因演算法；判斷式調整；機械式調整

A Knowledge Guided Adaptive GA based Forecasting Model in Consumer Packaged Goods Industry

Student: Chin-Lien Wang Advisor: Dr. Li-Chih Wang

Department of Industrial Engineering and Enterprise Information,
Tunhai University

ABSTRACT

Due to severe competition, price discount and non-price promotion campaigns are very commonly used by consumer packaged goods (CPG) manufacturer to retain competitiveness. However, dramatic change in sales volume in promotion session and sparse observations like specific promotion and lunar holidays' sales in the dataset make statistical models struggle, even though in general the latter are more reliable in forecasting performance, most of them can't use contextual information which is exploited in judgmental forecasting, a very common practice in CPG industry, however, the latter is subject to various kinds of biases and inconsistency.

Decomposition regression seems to be the natural option to integrate both methods in that it can incorporate contextual factors like promotion effects and holiday effects into the model, by "divide and conquer" it is capable of alleviating the information processing overload of forecasters. However, because these contextual factors usually are special events, related historical data are sporadic and don't have substantial variations, causing collinearity to arise, it becomes difficult for least square estimators to have a proper estimation against parameters.

In this thesis, a domain knowledge guided genetic algorithm (KGAGA) in the first stage is employed to address this issue by using MAPE as fitness function, instead of more commonly used MSE, to evaluate each candidate solution and alleviate the impact of square operations of outliers. Besides, a set of parameter constraints are set up based on contextual knowledge to ensure these parameters derived are truly meaningful and reflective to the real world. In particular, a detect and escape mutation algorithm (DEMA) is employed to detect any local pitfall (suboptimum) with a simple moving average metric, thereafter a loop of combination of broad search with ratio and deep search mutation operators to dramatically increase population diversity until the pitfall has been escaped. The crossover operator in which each parent to mate is selected randomly from a different group of different fitness may partially solve the dilemma of selection of fittest which is the ultimate cause of premature convergence. KGAGA competes favorably with ordinary GA (OGA) in efficiency and significantly outperforms the latter in effectiveness with its multiple-reconverging capability.

In case there are anticipated variations in the forecasting horizon which can't be handled by the regression model alone, a mechanical adjusting mechanism, formulated in a set of supplemental equations encompassing lunar-new-year seasonal index realignment, proportional adjustment of mixed effect of promotions and holidays at forecasting horizon in the second stage, coupled with the consideration of the weekend effect, can be used to deal with anticipated time shifting problem and reassess mixed effect of these contextual factors consistently and effectively without subjective judgments in judgmental adjustment to avoid possible bias and inconsistency therein.

Empirical results of a channel retailer of a CPG brand manufacturer in Taiwan reveal that KGAGA can be a better alternative than least square estimators in parameter estimation and multi-period out-of-sample forecasting considering contextual factors. It also beats OGA, ARIMA, exponential smoothing, and NAÏVE model in multi-step out-of-sample forecasting. Besides, the proposed mechanical adjustment mechanism could significantly reduce MAPE of weekly forecasts across seasons with improvement mainly coming from large size combined adjustments.

Keywords: Contextual information; Forecasting model; Genetic algorithms; Judgmental adjustment; Mechanical adjustment

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Background

According to Hoover's online, consumer packaged goods (CPG) industry is composed of companies that design, manufacture, and/or market apparel, cleaning products, hand and power tools, home furniture, house-wares, sporting goods, linens, and consumer electronics and appliances. Consumer packaged goods is a type of goods that is consumed every day by the average consumer. The goods that comprise this category are ones that need to be replaced frequently, compared to those that are usable for extended periods of time. While CPG represent a market that will always have consumers, it is highly competitive due to high market saturation and low consumer switching costs.

In this market, when shortage occurs, about 20% of the customers leave the store without buying any goods. And about 15% of promoted volume is lost because of stock-outs, see Cooper et al. (1999). Hence, availability of products in the stores is a critical factor for promoted sales, and most CPG manufacturers tend to manufacture in a make to stock style, and forecasting accuracy becomes a critical factor influencing the cost and therefore the profit of the company. Accurate forecasting can significantly reduce both the out-of-stock cost and overstock cost leading to greater company net profit and higher customer satisfaction. Thus, demand forecasting becomes a critical task in CPG industry.

Due to severe competition and high consumer sensitivity to price in CPG industry over the past three decades, promotion has become increasingly a critical factor of success, and sales are highly related to promotional activities. To properly forecast unit sales of a company's products in this industry, it is imperative for forecasters to take this contextual factor into account. Besides, traditional lunar holidays such as Dragon Festival, Tomb Sweeping, Moon Festival, and Lunar New Year (LNY) all have an obvious and positive effect on unit sales of CPG products, particularly the dates of traditional lunar holidays are moving across years, check Figure 1.1 below, it's quite intuitive to consider holiday effect as the second contextual factor of sales forecasting in CPG industry.

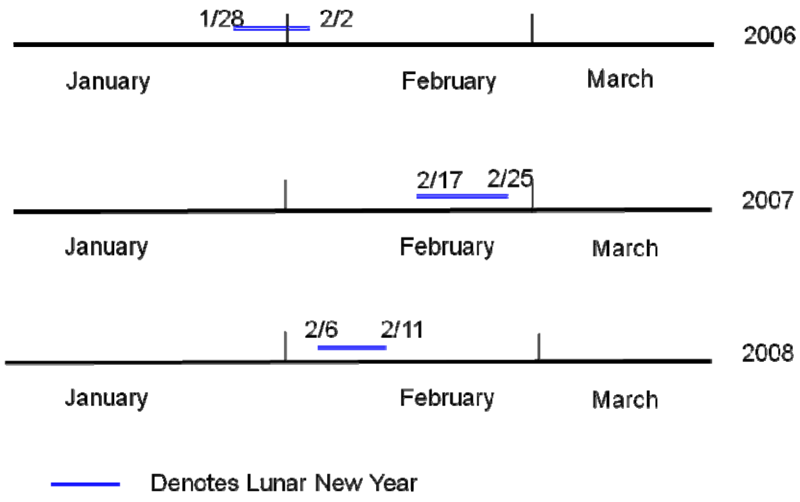

Figure 1.1.A typical moving holiday like LNY in ordinary calendar across years.

1.2 Motivation

Most statistical models can generate consistent forecasting results but are usually devoid of the flexibility and comprehensiveness of contextual information employed at judgmental forecasting, predictors and users of forecasts from the latter, however, are often tormented with the issue of inconsistency due to bias involved, which was pointed out clearly in Sanders and Ritzman (1992), Armstrong and Collopy (1998), and De Gooijer and Hyndman (2006).

From practical point of view, regression is the best choice to integrate both kinds of methods mentioned above (Edmunson, 1990; Bunn and Wright, 1991; Armstrong et al., 2005), because this approach is able to incorporate critical contextual factors in the model and gets consistent results in estimation and forecasting. In regression modeling, the conventional ordinary least square (OLS) still is one of the most widely used estimators to identify significant factors and estimate model parameters in linear regression (Draper and Smith, 1998; Rawlings et al., 1998). However, it suffers from limitations posed by the presence of outliers (Cook, 1977; Rawlings et al., pp. 330-331, 1998; Meloun and Militky, 2001), relatively small sample size (Belsley et al., 1980; Belsley, 1982; Yu, 2000), and multi-collinearity.

Multi-collinearity is the condition of one predictor variable which can be expressed as the exact or near linear combination of other predictor variables

(Gunst and Mason, 1977). Predictor variables with multi-collinearity contain much less information regarding their relationship with dependent variable. As pointed out by Smith and Campbell (1980), collinearity is ultimately caused by insufficient variation of observations related to predictor variables in the dataset. The overlapping or redundant of predictable variables causes inflated variance of variable coefficients, as a result, truly critical variables to become insignificant, or to have incorrect value or sign in parameters estimated, see Slinker (1985). These mis-specified model of inadequate variables and incorrect parameters, make accurate demand forecasting in CPG industry almost a mission impossible.

To tackle collinearity issues mentioned above, lots of studies in the literature have introduced remedies like ridge regression, partial least square, principle component regression analysis and so on with different degree of success but not without compromise, see Smith and Campbell (1980), and Yu (2000).

On the other hand, there are also some investigations of this topic using population based heuristics like genetic algorithm (GA), which keeps a population of solutions in terms of bit strings (chromosomes) evolving through crossover and mutation to better solutions from generation to generation with the selection principle of survival of the fittest. With its parallel search capability, GA is able to derive acceptably good solution of parameters in parameter estimation efficiently even with a small sample compared to other alternatives. Coupled with proper parameter constraints and fitness function, GA can give very promising results, check Wang and Wang (2009).

However, ordinary GA (OGA) in general has the problem of premature convergence and unstable performance. To make an effective search for solutions in the search space during a GA run, GA must keep a good balance between its exploring function, performed mostly by crossover operator, and exploiting capability, performed mostly by mutation operator, to make an adequate search. However, the exploring function of OGA usually is depleted due to the fact that after a couple of tens of iterations of crossover operations, the genetic material in the population become increasingly homogeneous, causing OGA hard to make any further exploration in the search space, and the search will usually end up getting stagnation in the local optimum.

Therefore, the control or maintaining of population diversity has become an important issue to circumvent the local optimum in the literature (Patrakis and Kehagias, 1998; Morrison and De jong, 2002; Ursem, 2002). For this purpose, a variety of diversity measures have been proposed. For example, Ursem (2002) used a distance-to-average-point diversity measure. He kept an upper threshold and a lower threshold for diversity measure. As long as diversity measure is above the upper threshold, control mechanism switches to diversity-decreasing operator. On the other hand, as diversity measure drops below the lower threshold, control mechanism switches to diversity-increasing operator until diversity measure reaches the upper threshold. He claimed that this method outperformed other algorithms on test problems. However, the time computation complexity of his diversity measure is O(l*n2), and hinders its widespread application.

Forecasting, most often, is subject to a variety of uncertainties in the forecast horizon, among them, random variations are unknown and hence no way to reliably deal with in advance, nevertheless, anticipated variations in promotion (Kumar and Pereira, 1997), and in holiday effects (Liu, 1980) usually can be expected before hand. If these anticipated variations are considered and incorporated into the model and finally reflected in the forecasts properly, forecasting accuracy usually could be significantly improved.

In CPG industry, judgmental adjustment is often used to adjust for promotion lift (Heerde et al, 2000) and post-promotion dip (Mace and Neslin, 2004) to improve forecasting accuracy of promotional sales. Most researchers in judgmental adjustment agree that, if the contextual information used is reliable, the performance of judgmental adjustments will be better than that of judgmental adjustments using unreliable one (Sanders and Ritzman, 1992; Remus, O’Conor, and Griggs, 1995; Lim and O’Conor, 1996). Yet judgmental adjustments still have all kinds of bias like cognitive bias, double-counting bias, political bias, and so on inherent in judgmental forecasting and hence still have the issue of inconsistency (Tversky and Kahneman, 1974; Armstrong, 1985; Sanders and Manrodt, 1994).

1.3 Methods

As mentioned above, regression model fitting with conventional estimator like OLS usually is tormented with multi-collinearity generating incorrect, unexplainable, or even meaningless parameters. However, model fitting with alternative estimator like GA, we must face the issue of premature convergence and performance instability. In this thesis to tackle these issues, we start with a log-linear regression model incorporating price, and a group of non-price promotion related dummy variables (Kumar and Pereira, 1997; Heerde et al., 2002a; 2002b) and decay factor, all of them are decomposed and assessed with a quantitative method called knowledge guided adaptive genetic algorithm (KGAGA) in the first stage:

1. Knowledge guided genetic algorithm (KGGA)

   The objective of KGAGA resides in deriving optimal solution or near optimal solutions of forecasting model parameters. To prevent the pitfall of premature convergence in OGA and achieve our objective, we use a novel version of GA called knowledge guided genetic algorithm (KGGA). In the literature, there are a variety of ways to incorporate domain knowledge into GA.

   White and Yen (2004) recommended the use of domain knowledge embedded in an improved crossover through ant pheromone memory to reduce the computation complexity of traveling salesman problems. Xie and Xing (1998) emphasized the importance of incorporating domain-specific knowledge in representation scheme and local search algorithms of general structure evolutionary algorithms. Schoenauer and Sebag (2002) pointed out that domain knowledge can be used to restrict the search space by a constrained Genetic Programming. Besides, they also argued that domain knowledge can also help representation-independent operators. For instance, the crossover operator used for the Voronoi representation does outperform the blind exchange of Voronoi sites.

   In this study, domain knowledge is used in two facets of KGAGA:

   (1)The constraints for model parameters: Leeflang and Wittink (2000) emphasized that robustness can be achieved with a model structure constraining answers to a meaningful range of values reflecting what is in the real world. If the actual values of a cue factor are constrained, the model counterpart should satisfy the same constraints. Therefore in this study, model parameters like price elasticity, and effects of various promotion mixes, as well as decay rate in Eq. (1) are explicitly defined in a constraint set in the decoding procedure of KGAGA according to findings in the literature and life experience about the product, as a result, no punishment function related to violation of constraints is needed, and a more realistic set of parameters, which are more meaningful and easier for interpretation, can be derived.

(2)A fitness function of mean absolute percentage error (MAPE), which, without the square operation of errors, as James and Stein (1961) exclaimed, an estimator which under squared error loss dominates the least squares estimator, and a pragmatic constraint on coefficient of variables which are set initially according to findings in the literature, and go through an iterative calibration process, we believe will, to a certain extent, alleviate the issue of inflated influence of outliers in parameter estimation by OLS and problems caused by collinearity.

2. A detect and escape mutation algorithm (DEMA).

In multimodal optimization, problem like continuous parameter optimization tackled in this paper, there are many suboptimal solutions in the search space (Miller and Shaw, 1996; Singh and Deb, 2006). As the search of solution via evolutionary algorithm like OGA goes on, the number of consecutive repetition of currently the best solution usually gradually increases with time, signifying a decreasing population diversity, or worse, a local pitfall arising, eventually the search will converge to a local optimum prematurely. In the literature, there are a large variety of methods trying to overcome this problem. Among them, diversity control via parameter adaptation looks promising, however, most of them struggle because of the time complexity of diversity measure like Hamming distance and information entropy incurred (Patrakis and Kehagias, 1998; Morrison and De jong, 2002; Ursem, 2002).

DEMA, which will be described and explained in more detail in section 3, is employed in this study to detect any suboptimal pitfall in the search space with a moving average of fitness improvement (*MAFI(l)*) in consecutive best solutions currently of recent *l* generations within KGAGA. Once repetition of *l* same best solution occurs, *MAFI(l)* will be zero, signifying the pitfall arising. As soon as a pitfall is found, a loop of combination of broad search operator with a predefined probability and deep search operator is used to dramatically increase the population diversity to escape the pitfall and trigger a new converging process. Contrary to most OGA having only one major convergence process, DEMA enables KGAGA to re-converge multiple times in the course of its search, the best solution found so far gets improved from convergence to convergence, as a consequence, the chance to get optimal or near optimal solutions is increased tremendously. Moreover, the computation complexity of *MAFI(l)* is a minor O(2) per generation based on fitness obtained, which is considerably simpler than typical $O(l*n^2)$ of Hamming distance or information entropy (Morrison and De Jong, 2002; Ursem, 2002) alike, thus, the extra heavy cost of computation can be saved.

3. A diversity maintaining crossover operator.

   Before selection of mates, the population is reordered according to ranking of fitness, followed by the division of the population into two groups with roughly equal amount of different fitness values. Each parent to breed is randomly selected from a different group. This method ensures the diversity of fitness value in each pair of parents and therefore diversity of gene material between two parents selected and, as a natural result, the diversity of offspring.

After parameter estimation with KGAGA, every parameter of regression model is derived, the re-composition of parameters can be used for forecasting promotional sales without adjustment—point forecast. As interval forecast is concerned, the program of KGAGA can be adjusted to derive MSE of estimation error in addition to the fitness function of MAPE. Thus, the original point forecast added or subtracted square root of MSE multiplied by a confidence level z* value (90%, 1.645; 95%, 1.96,…, and so on), an interval forecast can be thus derived. Therefore KGAGA can be used for forecasting purpose in the form of both point forecast and interval forecast.

In the second stage, an adjustment mechanism capable of handling and reflecting, without any subjective judgment, detailed changes anticipated in the forecasting horizon yet could not be handled by the regression model alone is introduced. This adjustment mechanism, a natural extension of the model, consisting of seasonal index realignment and proportional adjustment in a set of equations, is able to make appropriate adjustments consistently and effectively improving the forecasting accuracy of initial forecasts of the model compared favorably to other popular alternatives like Box-Jenkins ARIMA (Box and Jenkins, 1994; De Gooijer and Hyndman, 2006) and exponential smoothing (Gardner, 1985, 2006; Taylor, 2003 ; Armstrong, 1984) as well as Naïve model with adjustment.

Usually in the same period of time in lunar calendar between different years, the feature of promotion activities held is very similar to each other, in this thesis, this makes the nearest neighbor method an ideal underlying principle of mechanical adjustment of forecasting in our thesis. Generally speaking, the effect of each promotion mix of promotion sessions in last year is estimated with KGAGA based on historical data, the same or the most similar effect of promotion mix is used to forecast next year's sales in promotion session within forecasting horizon.

The weekend effect is also a critical consideration in our mechanical adjusting mechanism, because in weekly sales of CPG products, the unit sales in the weekend usually accounts for about half or more of a whole week's unit sales, see Singh and Takhtani (2005). To more precisely forecast weekly promotional sales in promotion session, it's imperative to take this factor into account.

In addition, moving holidays, mainly composed of lunar holidays, are also very important in our mechanical adjustment mechanism, due to the fact that there are only seven days a week, if in a specific week there is a moving holiday which keeps moving from a year to another year in its arising timing, in particular, if the moving holiday arising the weekend in the forecasting horizon rather than arose in the weekday in the training period, it does make sense to adjust forecast of weekly unit sales according to specific change of holiday condition based on holiday effect assessed from training dataset. Thus, the combined effect of promotion effect and holiday effect as well as weekend effect can be adjusted more correctly for anticipated variation in the forecasting horizon according to parameters estimated based on historical data in training period.

1.4 Objectives

This study bears many folds of objectives which can be listed below:

1. Proposing an adequate model estimator—a knowledge guided adaptive genetic algorithm (KGAGA) in regression modeling, particularly in the presence of collinearity, through a comparative study of LS estimator and OGA in both in-sample parameters estimation and out-of-sample forecasting, respectively. In addition, empirical results of weekly out-of-sample forecasting from ARIMA, ES and NAÏVE are also compared.
2. Estimating properly the effect of various kinds of promotion mixes applied at a typical company like company A in this study and providing managerial implications.
3. Presenting a mechanical adjustment mechanism to significantly improve initial forecasts of the regression model estimated via KGAGA in this research without judgment regarding forthcoming promotion events and holidays.
4. Verifying the existence of weekend effect and holiday effect on sales. Day of the week effect shows its appearance in the literature of finance, environmental science, and so on. But this effect seldom formally addressed in marketing, especially in promotion. From our everyday life experience, we know for sure about its existence. This study will incorporate holiday effect in our model, assess its effect, and take the weekend effect, holiday effect, and promotion effect into account in assessing mixed effect of promotions in mixed period, which has two kinds of promotions or two kinds of holiday effects arising in a single week, as well.

1.5 Framework of this thesis

In this thesis, section 1 briefly delineates the settings, the motives, the methods, and the objectives of this study. Section 2 gives some descriptions about the promotion intensive characteristics of CPG industry, the origin and extension of related models as well as some fundamental principles and ideas and methods this thesis used in the literature. Section 3 introduces the contextual forecasting model without holiday effects and justifies why the author uses KGAGA to estimate parameters of the model instead of OLS and

OGA in terms of MAPE, and Hamming Distance Ratio (HDR). Section 4 is part 1 of the empirical study of the contextual forecasting model without holiday effects, in which the background and design of this empirical study is portrayed, in addition to the model fitting, an emphasis is put on mode checking in which the test of collinearity and autocorrelation is focused, besides, a comparative tabular analysis of various forecasting methods without adjustments.

Section 5 briefly describes the contextual forecasting model, however, the emphasis is put on the mechanism of mechanical adjustments comprising seasonal index realignment (SIR), proportional adjustment (PA), and the combination of the former two called total adjustment (TA).

Section 6 is part 2 of the empirical study of the contextual forecasting model with holiday effects with a focus on forecast adjustments. Section 7 concludes this thesis. The framework of this thesis can be portrayed in a flowchart like Figure 1.2.

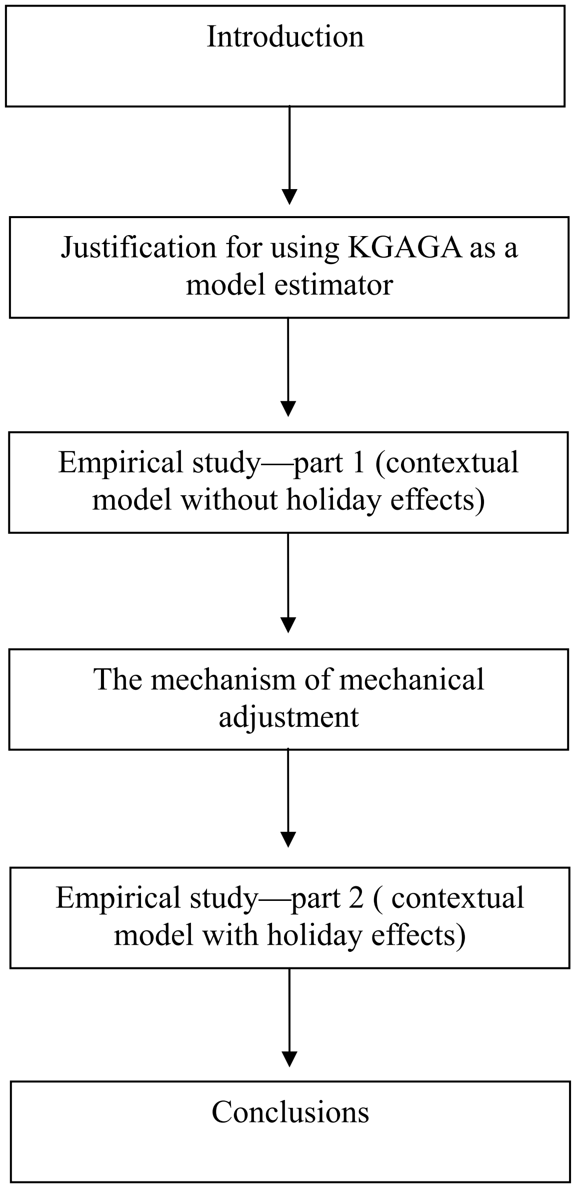

Figure 1.2. Flowchart of this thesis.

2. Literature Review

2.1 The Underlying Context of CPG industry

Over the past three decades, CPG manufacturers increasingly employ retail promotions to fight increasing competition and losing market shares for their own brands at the retail level. The objectives of manufacturers for retail promotions vary, but mainly centered around being competitive, by motivating increased sales through brand switching, increased consumption, and purchase acceleration. Manufacturers would like to know the effect of retailer promotion (type, timing, and frequency of promotions, and magnitude of price cuts, and so on) compared to that of other brands, see Kumar and Pereira in (1997).

Tellis and Zufryden (1995) examined optimal retail promotion decision concerning which brands to offer, how deep the price discounts should be, when they should be offered using a mathematical programming model approach. Very often, manufacturers offer trade promotion support to stimulate increased sales for their own brands. Such support may be performed through temporary price discount, see Duvvuri, Sri Devi et al. (2007), and Hardesty and Bearden (2003); local advertising or features, point-of-purchase displays, see Allenby and Ginter (1995) for a reference; free gift, see Raghubir (2004). Other tools like sweepstakes/drawings, and contest, and so on are common also, see Narayana and Raju (1985).

Due to various environmental and technological developments, today's retailers, generally interested in category sales and profits, have grown to a position of power, enabling them to be selective in passing through manufacturers' trade incentives in the form of trade promotions.

2.2 Recent Related Researches in Promotional Sales Forecasting

Cooper et al. (1999) introduced a promotion planning system called PromoCast. It used a traditional market response model in terms of a loglinear regression similar to that in SCAN*PRO developed to quantify the effects of promotions carried out by retailers for individual brands (Wittink et al., 1988). The model of SCAN*PRO can be formulated below in Eq. (2.1). This regression model aims to extract information in 67 variables from past promotion dataset for each stock keeping unit (SKU) in each store within a retailer chain, but with inherent limitations of historical promotion data pertained to a specific item and to a specific store without even considering competitor's actions.

$$Q_{kjt} = \left[\prod_{r=1}^{n} \left(\frac{p_{krt}}{\overline{p}_{krt}} \right)^{\beta_{rj}} \prod_{l=1}^{3} \gamma_{lrj}{}^{D_{lkrt}} \right] \left[\prod_{t=1}^{T} \delta_{jt}{}^{X_t} \right] \left[\prod_{k=1}^{K} \lambda_{kj}{}^{Z_k} \right] \ell^{\varepsilon_{kjt}} \tag{2.1}$$

Where,

Q_{kjt} denotes unit sales of brand j in store k, week t;

p_{krt} denotes unit price for brand r in store k, week t;

$\overline{p}_{krt}$ denotes regular unit price for brand r in store k, week t;

D_{1krt} denotes an indicator variable for feature advertising: 1 if brand r is featured but not displayed by store k, in week t; 0 otherwise;

D_{2krt} denotes an indicator variable for display: 1 if brand r is displayed but not featured by store k, in week t; 0 otherwise;

D_{3krt} denotes an indicator variable for the simultaneous use of feature and display: 1 if brand r is featured and displayed; 0 otherwise;

X_t denotes an indicator variable: 1 if observation is in week t;

Z_k denotes an indicator variable: 1 if observation is in store k;

β_{rj} denotes the own-brand (r = j) and cross-brand ($r \neq j$) price discount elasticity;

γ_{lrj} denotes feature only (l = 1), display only (l = 2), feature & display (l = 3) multipliers;

δ_{jt} denotes multiplier for brand j, week t;

λ_{kj} denotes store multiplier for brand j, store k;

ε_{kjt} denotes the disturbance term.

However, Fokens et al. (1999) proposed a dynamic version of SCAN*PRO model, and can be formulated below:

$$S_{ik,t} = \lambda_{ik,t}\delta_{k,t}\left(\frac{P_{ik,t}}{\overline{P}_{ik,t}}\right)^{\theta_{ik,t}} \prod_{l=1}^{3} \mu_{lik,t}{}^{D_{lik,t}} \prod_{j \neq k}^{J} \left[\left(\frac{P_{ij,t}}{\overline{P}_{ij,t}}\right)^{\theta_{j,k}} \prod_{l=1}^{3} \mu_{ljk}{}^{D_{lij,t}} \right] \ell^{\mu_{ik,t}} \tag{2.2}$$

Cooper and Giuffrida (2000) supplemented the PromoCast system, see Cooper et al. (1999), with data mining techniques to extract information from the many-valued nominal variables, which may be more specific to a particular retailer and geographic area, to conduct necessary adjustment against the forecast made with market-response model. It is based on rules inducted from the residuals which are not possible for traditional statistical model owing to hundreds of effects of dummy variables to be incorporated and estimated. They claimed that by use of this data-miner, case error could be reduced by 9% across all promotion events.

Trusov, Bodapati, and Cooper (2006) tackled the limitations inherent to the data mining techniques employed in PromoCast system by using a new data mining algorithm. It allows for set-valued features in the rule syntax, the size of the rule base has thus been significantly reduced, and the forecasting performance also could be improved as much as 50%.

Lee, Goodwin, and Fildes (2005) used a laboratory experiment to simulate a sales forecasting task undertaken by manufacturers distributing products to supermarkets running promotion campaigns. It aims to test hypotheses concerning the validity of increasing levels of support to derive more accurate forecasts by 54 students employing a computerized forecasting support system. Experiment results show that a simple support system can significantly improve the accuracy of forecasts under some conditions.

Gur et al. (2009a) experimented with 30 forecasting models and concluded that regression tree with features constructed from promotional sales time series data outperformed other forecasting methods. It provided as much as 65% improvement over the forecasting performance of the benchmark model using ES with lift adjustment for promotion in SKU-store level sales from 4 stores of a grocery retailer in Europe. The bench model can be formulated below:

$$\hat{D}_{ist} = \begin{cases} M_{ist}, if\ Disc_{ist} = 0, TV_{it} = 0, Radio_{it} = 0, Window_{it} = 0 \\ M_{ist} + \hat{L}_{ist}, otherwise \end{cases} \tag{2.3}$$

and

$$M_{ist} = \begin{cases} (1-\alpha)M_{is(t-1)} + \alpha S_{is(t-1)}, if\ Disc_{ist} = 0, TV_{it} = 0, \\ \quad Radio_{it} = 0, Window_{it} = 0 \\ M_{is(t-1)}, otherwise \end{cases} \tag{2.4}$$

Gur (2009b) proposed a driver-moderators method exploiting domain knowledge to define features reflecting potential drivers of sales in promotion and potential moderators of these drivers' effects on sales with the employment of an epsilon insensitive support vector regression with L1 norm regularization to simultaneously select the relevant ones among the 600+ features and to estimate the model parameters. Its forecasting performance in terms of MAE is said to be better than regression tree with extensive features, see Gur et al. (2009a), in one-step out-of-sample forecasting of daily sales of 155 items of black tea in 5 stores in Turkey.

2.3 Moving Holiday Effects

As Lin and Liu (2003) remarked, consumption, production, and other economic behavior in countries with large Chinese population including Taiwan are strongly affected by these lunar holidays. For example, production accelerates before lunar new year, almost completely stops during the holidays and gradually rises to an average level after the holidays. This moving holiday often creates difficulty for empirical modeling using monthly data and they employ an approach that uses regressors for each holiday to distinguish effects before, during and after holiday.

Soukup and Findley (2000) also mentioned that when the date of a holiday shifts from year to year, the effect of the holiday can influence two or more months in a way that depends on the date. If such holiday effects are ignored, then models fit to the time series will often have poorer forecasting ability.

Liu (1980) analyses the effect of the lunar calendar-based Chinese New Year holidays on Taiwan highway traffic volume between 1963 and 1976. He uses linear regression in conjunction with an ARIMA and concludes that sample patterns may be sufficiently disrupted by calendar structure that adjustment is warranted.

2.4 Use of Contextual Knowledge

Edmundson et al. (1998), and Sanders and Ritzman (1992) found that contextual knowledge of the forecast variable was a significant contributor to accuracy in product forecasting. This was especially true for products of importance to the company. Intimate knowledge of the underlying forces shaping the product (e.g. market plans and competitor issues) is crucial in deciding about future direction of the forecast variable.

Information and knowledge related to sales, usually are possessed by personnel across divisions within a company, must be collected and used in sales forecasting in order to improve its accuracy. Sanders and Ritzman (1992), Lim and O'Connor (1996), and others emphasize the importance of contextual knowledge such as product knowledge, weather information and so on to improve the accuracy of judgmental forecasting. Other researchers like Donselaar et. al. (2001), and Thonemann (2002), as well as Abuizam and Thomopoulos (2005), and others advocate the use of advance demand information (ADI) to assist forecasting.

In a contrast to most other forecasting methods, decomposition regression model can be benefited from the use of domain knowledge. People made more accurate forecasts than those made by either judgmental or quantitative method alone, see Armstrong et al. (1993). Lopes (1983) suggested that extrapolative methods would be strengthened by making provisions for causal or explanatory information to be used when such information is available.

New marketing initiatives, promotion plans, actions of competitors, industry developments, manufacturing problems and other forms of contextual information dominated discussion in each of the forecasting meetings we attended. These forecasting meetings always involved discussions of a number of variables or cues, not just one. It is possible that the task of incorporating new contextual information into the final forecast is upset by the tendency to over-inflate the influence of the past contextual information on the time series, and thus to overreact to the last actual value. Handzic(1997) examined the ability of people to utilize up to three cues in addition to the time series. They were, at best, able to utilize two cues. This suggests that there may be a severe information overload problem, see Lawrence et al. (2000).

Cue information, see K. Nikolopoulos et. al. (2007), of promotion mixes of a specific product in the forecasting horizon, specified in a promotion proposal set up more than three months earlier, if adopted properly, may shed some light on sales forecasting. Also, forthcoming holidays can be checked in the calendar before hand.

As Lawrence et al. (2006) put it, the total set of data useful for forecasting are made up of two classes; the history data and the domain or contextual data. The history data are the history of the sales of the product. The domain data are in effect all the other data which may be called on to help understand the past and to project the future. This includes past and future promotional plans, competitor data, manufacturing data and macroeconomic forecast data.

2.5 Use of Multiple Regression

Multiple linear regression (MLR) is a common choice of method when forecasts are required and where data on several relevant independent variables (or cues) is available. The technique has been used to produce forecasts in a wide range of areas (e.g. Burger et al., 2001; Sadownik and Barbosa,1999) and there is evidence that it is often used by companies to derive forecasts of demand from marketing variables and various macroeconomic measures (e.g. Mentzer and Bienstock,1998).

Because they can include policy variables (such as the price of a product), causal methods are useful for forecasting the effects of decisions in government and business. This is particularly true when one has good domain knowledge, accurate data, the causal variable has a strong effect on the dependent variable, and the causal variable will change substantially. (Armstrong, 2006).

Often forecasts based on the perfectly reliable regression model perform better than the original forecasts produced by the less than perfectly reliable human (Armstrong, 1985; Camerer, 1981).

2.5.1 Issues of Collinearity

In order to obtain reliable parameter estimates the number of observations made on each variable should be significantly greater than the number of variables. (Broadhurst et al., 1996), otherwise, it is very easy to have collinearity. As Armstrong (2006) point out, to use causal models, one must identify the dependent and causal variables, and then estimate the direction and size of the relationships. This requires much data in which there are substantial variations in each of the variables and the variations in the causal variable are independent of one another.

The least squares estimator performs poorly in the presence of collinearity. Collinearity arises when there exist near-linear dependencies among the explanatory variables, the explanatory variables often contain overlapping information and thus some degree of collinearity is likely to be expected. The least squares estimators of regression coefficients are still unbiased but their variances are large so they may be far from the true values. (Kejian Liu, 2003)

The presence of multi-collinearity in least squares regression can cause larger variances of parameters estimates which mean that the estimates of the parameters tend to be less precise. As a result, the model will have insignificant test and wide confidence interval. Therefore, the more the multi-collinearity, the less interpretable are the parameters. (Adnan et al, 2006)

2.5.2 Issue of Sparse Data

As Goodwin (2005) put it, unlike judgmental forecasters, however, statistical methods struggle when past data are scarce. So they have difficulties in handling special events or changes in the environment, such as promotion campaigns or new government policies.

Analysis of this "sporadic" data should be conducted at the "micro" level in order to extract what information is available in the data. E.g., a separate smaller level analysis might be conducted around sales in a regional market coinciding with a specific event or sponsorship. The analysis would give insights into the effect of this event – which would have implications on other or future events. This type of information is then collected whenever possible and used in the overall marketing mix decision making process. (Teasley)

Sparseness in data is defined by sample sizes that are small relative to the number of variables. In contingency table analysis, tables are called sparse when cell counts are small. When data are sparse, application of multivariate procedures of analysis can be risky. In particular, and in addition to the numerical problems discussed by Haberman and Agresti, multivariate procedures that rely on the assumption that data were drawn from multivariate normal populations may produce biased estimates. (Eye, 2006)

2.5.3 Parameters Estimation—Use of Adaptive GA (AGA)

There are many different kinds of parameter estimation methods, each may be appropriately applied to suit specific context. In general, the following methods like least square (LS) type methods, maximum likelihood (ML), and genetic algorithm (GA) are commonly used. As for the parameter estimation of multimodal functions, GA is considered to be an appropriate tool, see Montero et al. (2005), and Hati and Sengupta (2001).

Adaptive genetic algorithm (AGA) is an enhanced version of genetic algorithm (GA) which simulates Darwin biological evolution by selecting encoded individuals (chromosomes) in the population with higher fitness (via a fitness function) through stochastic crossover and mutation to generate population of individuals (reproduction) more fitted to the environment (better solutions) from generation to generation.

In estimating parameters of complicated multivariate nonlinear models, GA is generally considered to be better than other alternative such as nonlinear least square, maximum likelihood method, and so on, due to its parallel search capability (e.g. Pham & Karaboga, 1997; Hati and Sengupta, 2001; Montero et al., 2005).

With its peripherally designed mechanism, AGA enables parameters like crossover probability and mutation probability to vary with the number of generations processed to keep proper diversity of the population, as generic GA usually has the drawback of approaching homogenous population after a certain number of iterations, to circumvent getting stuck too early in local solutions in the process of its search for optimal solutions or near optimal solutions. (Schaffer et al.,1989; Pham & Karaboga,1997; and Eiben et al., 1999).

Libelli and Alba (2000) makes mutation of a function of fitness produces a more efficient search. In which, the least significant bits are more likely to be mutated in high-fitness chromosomes, thus improving their accuracy, whereas low-fitness chromosomes have an increased probability of mutation, enhancing their role in the search. In this way, the chance of disrupting a high-fitness chromosome is decreased and the exploratory role of low-fitness chromosomes is best exploited.

Liu et al. (2003) uses a disruptive selection, in which, the probability being selected is based on the ranking value of individual guarantees the maintaining of diversity in the population, besides, the probabilities of crossover and mutation are also adaptively varied depending on the ranking value of individuals. They claim that this type of GA is capable of considerably improving premature convergence and find the optimal solution efficiently. In GA research and applications, one of the most important concerns is premature convergence. This occurs when the population in a GA reaches a state such that the genetic operators can no longer produce offspring which outperform their parents, as observed by many authors. (Xu and Gao, 1997)

A critical issue in studying premature convergence is the identification of the convergence and the characterization of its extent. Srinivas and Patnaik, for example, used the difference between the average and maximum fitness value as a yardstick to measure premature convergence in GA and then varied the crossover and mutation probabilities adaptively according to the measurement. On the other hand, the term population diversity has been used in many papers to study premature convergence. It is widely recognized that the decrease of population diversity leads directly to premature convergence. (Xu and Gao, 1997)

To make an effective search for solutions in the search space during a GA run, GA must keep a good balance between its exploring function, performed mostly by crossover operator, and exploiting capability, performed mostly by mutation operator, to make an adequate search. However, the exploring function of GA usually is depleted due to the fact that after a couple of tens of crossover operations, the genetic material in the population become increasingly homogeneous, causing GA hard to make any further exploration in the search space, and the search will usually end up getting stagnation in the local optimum.

Various studies showed that a selection process controls the level of exploration or exploitation by varying its selection pressure. Directing an evolutionary process towards exploration or exploitation is also possible by population resizing: With bigger population size, the search space is explored more than with smaller population size. The mutation and crossover operators also adjust the power of exploration and exploitation by respectively tuning their mutation rate and crossover rate towards either aspect. (Liu et al., 2009)

When are exploration and exploitation controlled during an evolutionary process? There are also two remarks to address this topic: fitness-driven and diversity-driven approaches. For the fitness-driven approach, fitness values are applied to determine the appropriate time to leverage exploration or exploitation. For example, if the best fitness value is improved at a given generation or has not improved for a number of predefined generations, ProFIGA (Population Resizing on Fitness Improvement Genetic Algorithms) increases population size. Otherwise, the population size is decreased. The 1/5 success rule is also classified in this category: The ratio of better fitness individuals generated by the mutation operator is regarded as an adaptation criterion to update the new mutation rate towards more exploration or exploitation. For the diversity approach, diversity measures (e.g., standard deviation and Euclidean distance) control when to perform more exploration or exploitation. For example, the Diversity-Guided Evolutionary Algorithm (DGEA) uses a distance-to-average-pointmeasure to alternate between exploration and exploitation phases. (Liu et al., 2009)

Premature convergence, or the loss of diversity before a satisfactory solution is found, is a persistent problem in evolutionary optimization. This reflects the fundamental trade-off between exploration and exploitation, or between thoroughness and speed in evolutionary search. If selection is too weak, progress is slow and many generations are required to find a solution. On the other hand, if selection is too strong, the population rapidly loses diversity and may become stranded on a local suboptimum. (Pepper, 2010)

While the trade-off between improving performance and preserving diversity cannot be avoided, it can be ameliorated through the efficient use of variation. Diversity within a population acts as the fuel of the selection process: it is required for selection to act, but is itself consumed in the process. However, selection algorithms differ not only in speed, but also in “fuel efficiency”, or rate of improvement relative to loss of variation. (Pepper, 2010)

Ursem (2002) used a distance-to-average-point diversity measure. He kept an upper threshold and a lower threshold for diversity measure. As long as diversity measure is above the upper threshold, control mechanism switches to diversity-decreasing operator. On the other hand, as diversity measure drops below the lower threshold, control mechanism switches to diversity-increasing operator until diversity measure reaches upper threshold. He claimed that this method outperformed other algorithms on test problems. However, the time complexity of his diversity measure is O(l*n2), thus hinders its popular usage.

2.5.4 Use of Decomposition Methods

Webby, O'Connor and Edmundson (2005) showed that, when a time series was disturbed in some periods by several simultaneous special events, accuracy was greater when forecasters were required to make separate estimates for the effect of each event, rather than estimating the combined effects holistically. (Lawrence et al., 2006) If the components of a complex series can be forecast more accurately than the global series, it helps to decompose the problem by causal forces. (Armstrong, 2006).

In the literature, most of the promotion sales forecasting are conducted with decomposition methods which are mainly implemented through multiplicative models incorporating promotional considerations. The appealing rationale for decomposition is that it allows judgmental tasks to be divided into cognitively less demanding subtasks. Thus, it is argued, if the responses to the subtasks are more accurate and can serve as inputs to sound aggregation rules, then it is reasonable to expect that the recomposed forecasts should outperform holistic assessments. (Salo and Bunn, 1995).

Decomposition methods are designed to improve accuracy by splitting the judgmental task into a series smaller and cognitively less demanding tasks, and then, combining the resulting judgments. (Lawrence et al., 2006) Armstrong et al. (2005) point out that a time series could be effectively decomposed when (1) uncertainty is high (2) forecasters can use domain knowledge to decompose the problem such that different forces can be identified for two or more component series, (3) the causal forces imply trends that differ in direction, and (4) it is possible to obtain forecasts for each component that are more accurate than the forecast for the global series. They claim that for nine series in which the conditions are completely or partially met, the median absolute percentage error (MdAPE) is reduced by more than half.

2.6 The Drawback of Judgmental Forecasting

While judgment can play a valuable role in forecasting, it can also be subject to biases and inconsistencies arising from cognitive limitations, political influences or confusion between forecasts, targets and decisions. Although additional information could serve to improve the forecaster's understanding of the environmental conditions at the time of the forecast, it also increases the complexity of the forecasting task and may impose a cognitive burden on the forecaster that exceeds human information processing capacity. (Stewart and Lusk, 1994) Sanders and Manrodt (2003) pointed out that biases inherent in judgmental forecasting include optimism, wishfulthinking, lack of consistency, political manipulation, and overreacting to randomness. Forecasters also have limited information processing capacity, which often causes them to oversimplify problems, to produce biased judgments and to behave inconsistently (Goodwin, 2002).

2.7 The Use of Judgmental Adjustment and Its Drawbacks

As Mathews and Diamantopolous (1986, 1989, 1990) demonstrated, the revisions of statistically generated forecasts using relevant domain knowledge enabled greater final forecast accuracy. Judgmental adjustment is likely to be beneficial is where the forecaster has important domain knowledge that is not available to the statistical method, such as knowledge about a forthcoming sales promotional campaign. (Fildes, 1999; Mathews & Diamantopoulos, 1990)

Judgmental forecasters and statistical methods have their own strengths and weaknesses (Blattberg & Hoch,1990). This has led to the application methods that integrate the two types of forecasts. These include a simple averaging of independent judgmental and statistical forecasts (Blattberg & Hoch, 1990), bootstrap models of judgmental forecasters (O'Connor, Remus, & Lim, 2005), and statistical methods that are designed to remove systematic biases from judgmental forecasts (Goodwin, 2000).

The accuracy of causal adjustment was lower when causal information was less reliable. (Lim and O'Connor, 1996). More Forecasting accuracy can be gained by making the additional information the basis of adjustment to an extrapolation of the time series, referred to hereafter as the extrapolation/adjustment method. (Edmundson et al; 1988). The forecast adjustments reflect expert knowledge on important factors not in the models, or a correction for model misspecification. (Bunna and Salo, 1996).

Edmundson et al. (1988), Sanders and Ritzman (1992) demonstrated how non-time series causal information improved the accuracy of the final forecast over the statistical and judgmental initial estimates. Managers who possessed considerable product knowledge were also found to successfully adjust exponential smoothing forecasts (Mathews and Diamantopoulos, 1986, 1989).

There are a number of cognitive biases in relation to causal forecasting. Tversky and Kahneman (1980) defined four types of evidence (D) in conditional judgement P(X/D): causal, diagnostic, indicational and incidental. Schustack and Sternberg (1981) also found that causal inference was suboptimal and subject to many cognitive biases. They include the confirmatory trap, the underestimation of negative evidence, the neglect of base-rate information and insensitivity to the notion of sample sizes.

Statistical time series forecasting methods are designed to measure and extrapolate regular patterns in time series. However, in practice, many series have temporary discontinuities caused by sporadic events, such as promotion campaigns for a product. These events may be so infrequent that the ability of statistical methods to measure their effects is restricted by lack of data and they may therefore be treated as noise. This limitation of statistical methods is one reason for the widespread use of human judgment in business forecasting (Dalrymple, 1987; Kleinmutz, 1990; Sanders and Manrodt., 1994).

Despite the popularity of judgmental forecasting, research on the accuracy of judgment has suggested that it is subject to cognitive biases and inconsistency (Goodwin and Wright, 1994). Nevertheless, some research (Mathews and Diamantopoulos., 1990: Webby and O'Connor., 1996) has found that judgmental forecasters are able to make effective adjustments to statistical time series forecasts to take into account contextual information (i.e, any information in addition to that contained in the time series). It therefore seems reasonable to hypothesize that the appropriate approach to the problem of recasting discontinuous time series patterns is a combination of statistical methods and judgment, with the statistical forecast handling the regular time series pattern and the judge making adjustments to this in the light of sporadic events. (Goodwin and Fildes, 1999). Sporadic contextual information puts special demands on the judgmental forecaster and may therefore have a particular influence on the way that judgment is used. (Goodwin and Fildes, 1999).

Judgment can be valuable when the forecaster has access to important information about a forthcoming event that cannot be used in a statistical model

(Sanders & Ritzman, 2001). A typical event in sales forecasting would be a sales promotion campaign. Quantitative data on the effects of such campaigns might be scarce because of their infrequency or their diverse nature. Forecasters commonly search for past circumstances analogous to those which will prevail in the forecasting period in order to establish a basis for their judgmental forecasts. There are reasons to doubt the efficacy of an informal use of analogies. First, the forecaster may have to recall similar cases from memory and judge their similarity to the target case. Second, limitations in human information processing capacity may mean that the forecaster relies on a single recalled case. Finally, the forecaster will have to adapt the outcome of the past case to take into account the aspects of the target case that are different. (Lee et al., 2007) Judgmental revision of statistically generated forecasts, a common organizational practice, has received an equal amount of discouragement due to its potential to deteriorate accuracy. (Sanders and Manrodt, 2003).

3. Justification for Using KGAGA as an Estimator

3.1 Formulation of a Contextual Forecasting Model

The following equation is motivated by Dick R. Wittink *et al.,*'s dynamic market response model (Foekens *et al.,* 1999; Heerde *et al.,* 2002a; 2002b) and is modified to reflect the characteristics of promotional sales of CPG industry in Taiwan. The model can be formulated as

$$S_{it} = \lambda_i \left(P_{it} / \widehat{P}_i\right)^{\theta_i} \prod_{l=1}^{n} \mu_{l_i}{}^{D_{l\,it}} \beta_i{}^{E_{it}} \ell^{\varepsilon_{it}}, \forall t \in Q \tag{3.1}$$

where, i denotes an item number, i = 1,2,3….,I; t denotes specific number of period referenced, $1 \le t \le$ T. T is the total number of normal periods, and I is the total number of items involved.

Q denotes the set of referenced periods.

S_{it} is the total unit sales of the item i in period t under a retailer, for weekly sales, t actually represents a certain week in the referenced periods.

λ_i denotes the normal unit sales (baseline sales) of the item i without any promotion under a retailer.

$\widehat{P}_i$ is the list price of item i.

P_{it} is the discount price of item i during period t under a retailer, note that ($P_{it} / \widehat{P}_i$) is an observation of price discount ratio.

θ_i denotes the coefficient of price elasticity of item i under a retailer.

$D_{l\,it}$ is the l-th component of a vector of n indicator parameters of non-price promotion mix ($D_{1\,it}$, $D_{2\,it}$, ..., $D_{n\,it}$) of item i in period t. $D_{l\,it} = 1$ denotes a promotion mix of type l arises, since at a time only one condition arises, all other components in this vector have the value of 0. In this equation, n components are used to represent n+1 possible combinations of promotion activities, in which when every component has the value of 0, it signifies the condition of no promotion activity. The default value of $D_{l\,it}$ is 0.

$\mu_{l\,it}$ denotes the non-price promotion effect parameter (multiplier) of corresponding non-price promotion mix ($D_{l_{it}}$) of item i during normal period t under a retailer.

E_{it} denotes the indicator parameter of decay rate of price discount effect of item i in period t, $E_{it} = 1$ implies that there is a decay of price discount effect. Note that there is only one level of decay in this study, once the decay effect arises, it remains there for a certain price discount; whereas $E_{it} = 0$ signifies no decay of price discount effect.

β_i is the decay rate of price discount effect of item i, check Leung et al. (2008) in which they proposed consecutive promotion in a stochastic dynamic model using exponential decay function to deal with decay of promotion effects, also see Jedidi et al. (1999), and Burt (2000) for a reference.

$\ell^{\varepsilon_{it}}$ denotes an exponential function of residual error ε_{it} of item i in period t, ℓ is defined as the base of the natural logarithm.

From Eq. (3.1), we can see that the promotional sales of a certain item i in time t can be decomposed into the following multiplicative effect factors: normal sales (λ_i), price discount effect $\left(P_{it} / \widehat{P}_i\right)^{\theta_i}$, promotion mix effect $\prod_{l=1}^{n} \mu_{li}^{D_{lit}}$, and the decay effect ($\beta_i^{E_{it}}$) of price promotion, as well as residual error ($\ell^{\varepsilon_{it}}$).

Take natural logarithm in both sides of Eq. (3.1), we get the following linear equation in additive form:

$$\ln S_{it} = \ln \lambda_i + \theta_i \ln\left(P_{it} / \widehat{P}_i\right) + \sum_{l=1}^{n} D_{lit} \ln \mu_{li} + E_{it} \ln \beta_i + \varepsilon_{it}, \quad \forall t \in Q \tag{3.2}$$

Parameters in Eq. (3.2) in the following sections will be estimated by KGAGA, OLS, and OGA respectively. After that, parameters will be recomposed like Eq. (3.1) without the residual error term for forecasting purpose.

3.2 Parameter Estimation with KGAGA

In this study, we propose a knowledge guided adaptive genetic algorithm (KGAGA) to be used as a regression estimator rather than widely used least square estimators to estimate coefficients of variables under situations of small sample size, or model mainly composed of dummy variables, and sporadic variables which are usually regarded as insignificant by LS estimators owing to insufficient observations are available.

3.2.1 Issues of ordinary GA (OGA)

In estimating parameters of complicated multivariate nonlinear models, GA is generally considered to be better than other alternatives such as nonlinear least square, maximum likelihood estimation, and so on, due to its parallel search capability (Schaffer et al., 1989; Eiben and Michalewicz, 1999), even based on small size sample, it is still capable of deriving good results (Liu et al., 2003; Pham and Karaboga, 1997).

However, OGA often suffers from premature convergence owing to increasing homogeneity of genetic material in population during its search process driven mainly by selection of mate and fitness, and operators of fixed probability, leading mostly to a suboptimal solution, particularly in multimodal problems. Nevertheless, once in a while it will converge to a good solution by chance, the performance is unstable.

3.2.2 Features of KGAGA

Domain knowledge guided genetic algorithm (KGGA) and adaptive genetic algorithm (AGA) are two major features of KGAGA which will be described in the following subsections. In addition, the whole procedure of KGAGA will be delineated in great detail in the last part of this section.

1. Knowledge guided genetic algorithm (KGGA)

    The objective of KGAGA resides in deriving optimal solution or near optimal solutions of forecasting model parameters. To prevent the pitfall of premature convergence in OGA and achieve our objective, we use an algorithm called knowledge guided genetic algorithm (KGGA). In the literature, there are a variety of ways to incorporate domain knowledge into GA.

    White and Yen (2004) recommended the use of domain knowledge embedded in an improved crossover through ant pheromone memory to reduce the computation complexity of traveling salesman problems. Xie and Xing (1998) emphasized the importance of incorporating domain-specific knowledge in representation scheme and local search algorithms of general structure evolutionary algorithms. Schoenauer and Sebag (2002) pointed out that domain knowledge can be used to restrict the search space by a constrained Genetic Programming. Besides, they also argued that domain knowledge can also help representation-independent operators. For instance, the crossover operator used for the Voronoi representation does outperform the blind exchange of Voronoi sites.

In this study, domain knowledge is used in two facets of KGAGA:

(1)The mutation operator: Based on domain knowledge about the convergence process gained from the literature and feedback information derived from *MAIF(l)*, we adapt the scope of exploration and the size of target population to mutate. The former is performed via phase-varying mutation probabilities (p_m), with DEMA in the course of the search process. For example, in the normal phase, the mutation probability (p_m) is predefined to be a small one, about in the range of [0.008, 0.10] to search for better solutions. When a stuck is found, a loop of a combination of a broad (explorative) search mutation operator and a deep (exploitative) search mutation operator will be triggered. For the former, p_m will be dramatically increased, usually in the range of [0.3, 1] with multiple bits to be mutated to perform an explorative search to help KGAGA escape from local optimum. For the latter, the mutation probability is dramatically reduced to the range of [0, 0.001] to make an exploitative search while still keeping the capability to converge.

(2)The constraints for model parameters: Leeflang and Wittink (2000) emphasized that robustness can be achieved with a model structure constraining answers to a meaningful range of values reflecting what is in the real world. If the actual values of a cue factor are constrained, the model counterpart should satisfy the same constraints. Therefore in this study, model parameters like price elasticity, and effects of various promotion mixes, as well as decay rate in Eq. (1) are explicitly defined in a constraint set in the decoding procedure of KGAGA according to findings in the literature and life experience about the product, as a result, no punishment function related to violation of constraints is needed.

2. Adaptive genetic algorithm.

To make an effective search for optimal or optimal solutions in the search space during a GA run, GA must keep a good balance between its exploring function, performed mostly by crossover operator, and exploiting capability, performed mostly by mutation operator, to make an adequate search. However, the exploring function of GA usually is depleted due to the fact that after some iterations of crossover operations, the genetic material in the population become increasingly homogeneous, causing GA hard to make any further exploration in the search space, and the search will usually end up getting stagnation in the local optimum.

Therefore, the control or maintaining of population diversity has become an important issue to circumvent the local optimum in the literature. For this purpose, a variety of diversity measures have been proposed. Among them, Hamming distance and information entropy is two of the most widely used diversity measures in the literature.

He kept an upper threshold and a lower threshold for diversity measure. As long as diversity measure is above the upper threshold, control mechanism switches to diversity-decreasing operator. On the other hand, as diversity measure drops below the lower threshold, control mechanism switches to diversity-increasing operator. He claimed that this method outperformed other algorithms on test problems.

Hinterding *et al*. (1997) classified adaptation in evolutionary algorithm based on the mechanism of adaptation used in the search process into 2 types, namely static (constant parameters) and dynamic (parameters adapted with mechanism), which can be further classified into three subtypes: deterministic (value of parameters changed by deterministic rules without any feedback), adaptive (value of parameters changed with feedback) and self adaptive (parameters encoded onto chromosomes of individual to be adapted). Apparently, $MAFI(l)$ in this study is a feedback information for parameter adaptation, our adapting mechanism of mutation probability in KGAGA is belong to adaptive type.

Our mechanism of adapting parameters mainly focuses on adapting parameter of mutation operator embedded in DEMA. Here, DEMA enables mutation probability to vary in different phases of search process. Before getting stuck, KGAGA uses a normal search setting of ordinary mutation probability (p_m), once KGAGA gets stuck, KGAGA alternately employs a broad search mutation operator with a much higher p_m than ordinary mutation probability and a deep search mutation operator with setting of p_m lower than that of normal search, in order to restore the diversity of individuals in the population while maintaining its capability of convergence in the same time until the search gets out of the local optimum. Once KGAGA gets out of the pitfall, the control returns to the normal search. These p_m of different phase are control parameters to be tuned by experiments, however. Empirical results in section 6 prove that

3.3 The Whole Process of KGAGA

The whole process of KGAGA for model fitting is delineated in pseudo codes in Figure 3.1, whereas a more detailed description can be given below:

1. Initialization procedure

   The initial population is randomly created in the encoded form of a binary matrix, where there are *m* rows, each row of binary string in the matrix is an individual (candidate solution) which encompasses ν chromosomes, each chromosome, representing a parameter, is composed of γ genes, while each gene is represented by a binary code. Each individual consists of γ * ν bits of genes.

Procedure knowledge guided Adaptive genetic algorithm (KGAGA)
Begin
  Initialize a randomly created binary matrix as population.
  Decode chromosomes of each individual in population into real numbers with Eq. (3.4).
  Evaluate each individual based on fitness function of Eq. (3.5).

  While termination condition of program is not met do
    Select the best proportion of the population and reproduce.
    Sort the population based on ranking of fitness.
    Divide the population into two groups of equal number of different fitness values.
    Randomly select each parent from two different groups to breed via 2-point crossover operator to fill up the remains of next-generation population.
    Decode chromosomes of each individual in population into real numbers with Eq. (3.3).
    Evaluate each individual based on fitness function of Eq. (3.4).
    Sort the population newly generated based on rank of fitness.

Calculate mutate-pop-rate with Eq. (3.5).

Perform Detect and Escape mutation algorithm. Check the mutation procedure of section 3.3.

Decode chromosomes of each individual in population into real numbers with Eq. (3.3).

Evaluate each individual based on fitness function of Eq. (3.4).

Output the best solution of parameters in real numbers so far.

End of While

End of program

Figure 3.1. Pseudo codes of KGAGA

2. Decoding procedure

The binary string of individual is decoded back to a vector of real numbers standing for parameters of regression model. Note that each chromosome of an individual can be decoded onto a real number by the following equation:

$$RN = RN_{min}+(RN_{max}-RM_{min})*\text{accu}/(2^n-1) \tag{3.3}$$

Where, RN represents the decoded value (in real number) of a parameter such as price elasticity (θ_i), the effect of promotion campaign (μ_i), and decay effect of price reduction (β_i) in Eq. (3.1). RN_{max} and RN_{min} stands for the upper bound and lower bound of a constraint of parameter, respectively. While accu stands for original value (in real number) of binary string. And n stands for the length of binary string. Note that both RN_{min} and RN_{max} of the constraint are determined by domain knowledge and life experience. Each binary string is decoded in such a way that no constraint is violated, hence, no penalty function is needed.

3. Evaluation procedure

Each individual is evaluated by the fitness function using MAPE to alleviate the negative impact of outliers, thus increasing the reliability of parameter estimation, check Eq. (3.4).

Based on Eq. (3.2), the fitness function of KGAGA may be formulated as

$$FV_i = MAPE_i = (\sum_{t=1}^{T} \left| \ln S_{it} - \ln \hat{S}_{it} \right| / \ln S_{it}) / T \qquad , \forall t \in Q \tag{3.4}$$

Where, the term $\left| \ln S_{it} - \ln \hat{S}_{it} \right|$ is the absolute value of difference between natural logarithm of the actual sales volume ($\ln S_{it}$) of the *i*-th item and natural logarithm of the in-sample unit sales prediction ($\ln \hat{S}_{it}$) of the same item in period *t*. *T* denotes the number of normal periods. The objective of KGAGA as an estimator is to find a solution with the minimal $MAPE_i$. The smallest $MAPE_i$ found is updated once a smaller one is found in the solution search process.

4. Selection scheme

As for selection strategy is concerned, this study uses both elitism and diversity maintaining crossover operator. In each generation, the best individuals in terms of fitness occupying (1- mate_pop_rate) of the population are kept as elites which are sure to survive to the next generation. The remains of next generation population comprise offspring of parents selected randomly from two groups of equal number of different fitness values in the whole population ordered by ranking of fitness, each time only one pair is picked up to mate, the process doesn't terminate until the mate_pop_size of parents is met.

5. Crossover procedure

A control parameter called mate_pop_rate, which is employed to set the portion of population as parents to breed offspring. Mate_pop_rate also sets the upper bound of the portion of population to participate mutation. Investigations of Spears (1992); Kubalik and Lazansky (2000) showed that among one-point, 2-point, uniform and several other commonly used crossover operators for GA with large population, 2-point crossover operator is the least disruptive operator, therefore in this study, a 2-point crossover is used with modification.

The whole population is reordered with ranking of fitness first. Then, the population is divided into two groups of roughly equal amount of different fitness values. The pivot point is the mid point of fitness values in the whole population which is divided in such a way that the fitness value of one group is less than that of another group. After that, the operator randomly picks a parent from each group to ensure the diversity of genes in the couple selected to mate. The total number of pairs is about the half of mate_pop_size to breed offspring, each of them is a random recombination of genetic material from two parents to explore new promising areas in the search space.

6. Mutation procedure

The mutation operator aims at creating components of chromosome not included in the current population by changing the content of a position or multiple positions in the binary string of each chromosome of individuals in the population. After crossover operator is performed, the fitness of each individual in the population is evaluated and the whole population is reordered according to the ranking of fitness.

Similar to mate_pop_rate, a control parameter called mutate_pop_rate is used to set up the size of target group in the population to conduct mutation operation. The difference being that mate_pop_rate is deterministic before hand, whereas mutate_pop_rate is time-varying (increasing with generations), and can be determined with the following equation:

$$\text{mutate\_pop\_rate} = \text{mutate\_pop\_rate} + (\text{mate\_pop\_rate} - \text{mutate\_pop\_rate}) * (g / \text{generation}) \tag{3.5}$$

Where, g is current index of iterations run, mate_pop_rate is given in the beginning of KGAGA program, while generation is the predefined number of iterations to run. Eq. (3.5) means the mutate_pop_rate is directly increasing with the number of generations processed and reaches its upper bound which is mate_pop_rate eventually.

In this study, three kinds of mutation operators, namely normal_search, broad_search, and deep_search mutation operators, are used to exploit the best solution(s) found and to explore new promising areas also. Each operator is applied in a different condition of genetic evolution.

A measure called Moving Average of Fitness Improvement of l generations (MAFI(l)) is employed to detect the pitfall in the search process. MAFI(l) is the moving average of the improvement of best fitness value (MAPE) found so far of l generations, the algorithm to calculate MAFI(l) can be found in Figure 3.2. In ordinary generations without bottleneck or pitfall, normal_search mutation is used to make a small change in the composition of chromosome(s) (randomly picked one-position bit-flip in each chromosome with a small mutation probability (pm) to do a neighborhood search to improve the best solution(s) found so far.

From l + 1 generation onwards, KGAGA keeps a measure of pitfall called MAFI(l), once the MAFI(l) becomes zero, that is, there are l consecutive repetitions of the same best MAPE present and this implies a pitfall arising, a loop of combination of broad_search mutation operator and deep_search mutation operator, the former is executed with a predefined probability called ratio, is triggered, it won't stop until MAFI(l) becomes greater than 0 signifying the search escapes the pitfall in a broad sense. The broad_search mutation operator makes a free search with a much higher Pm than that of normal search with a predefined proportion of this combination and multiple positions are randomly picked in each chromosome of every individual in the whole target group of mutate_pop_size to conduct bit-flip. The deep_search mutation operator makes an exploitative search in the neighborhood of the best found solution so far with a much smaller Pm than that of a normal search. A single place is randomly chosen in each chromosome of every individual in the whole target group of mutate_pop_size to conduct bit-flip.

Procedure calculate *MAFI*(*l*)
    Begin
        If $g <= l + 1$and $g > 1$ and termination condition is not met then
            FI(g) = min_MAPE(g-1) – Min_MAPE(g).
            AFI = AFI + FI(g).
        If $g = l + 1$ then
            MAFI(g) = AFI(g) / l.
        Else if $g > l + 1$ then
            FI(g) = min_MAPE(g-1) – Min_MAPE(g).
            AFI = AFI + FI(g) – FI(g – l).
            MAFI(g) = AFI(g) / l.
        End of If.
End of procedure.

Figure 3.2. Pseudo codes of procedure calculate *MAFI*(*l*).

In the broad_search mutation operator, multiple positions of each chromosome of each individual in the group of individuals to be mutated are picked with the following probability:

$$mutate\_rate = mutate\_rate * i*(g / generation)^k \quad (3.6)$$

Where, g and generation are defined as earlier, while k is a control parameter usually set in the interval [1, 2], and i is the rank of each individual, based on fitness value, in the target group to be mutated. While mutate_rate in the right hand side of Eq. (3.6) is just a predefined mutation probability (pm). Eq. (3.6) implies that the fitter the individual is, the smaller the probability of the individual to have multiple positions in each chromosome to be mutated.

If after broad_search or deep_search mutation, the search gets out of stuck of local optimum (MAFI(l) becomes greater than 0), the mutation is returned to normal_search operator in which each individual is endowed with equal small probability to have one position in each of its chromosome get bit-flipped, see Skalak (1994), until another stagnation is found. Check Figure 3.3-3.4 for more detail about the algorithm of our adaptive mutation algorithm.

7. Termination of KGAGA

After evaluation of fitness of each individual, a mechanism is performed to check if the termination condition of the program is met, if the answer is yes, the program terminates, otherwise it goes to selection procedure.

```
Procedure Detect and Escape mutation algorithm
    Begin

        Calculate MAFI(l).
        If MAFI(l) = 0 and g > l then
            If rand < ratio of combination to conduct broad_search
                perform broad_search mutation operator.
            Else
                perform deep_search mutation operator.

        Else
                perform normal_search mutation operator.
        End of If
    End of algorithm
```

Figure 3.3. Pseudo codes of our Detect and Escape mutation algorithm (DEMA)

3.4. Analysis of KGAGA with Hamming Distance and Min MAPE

3.4.1 Hamming distance ratio

To facilitate the demonstration of detailed operation mechanism of KGAGA in terms of analytical graph of both Min MAPE and Hamming distance in this study, a measure called Hamming distance ratio (HDR) is formulated as following:

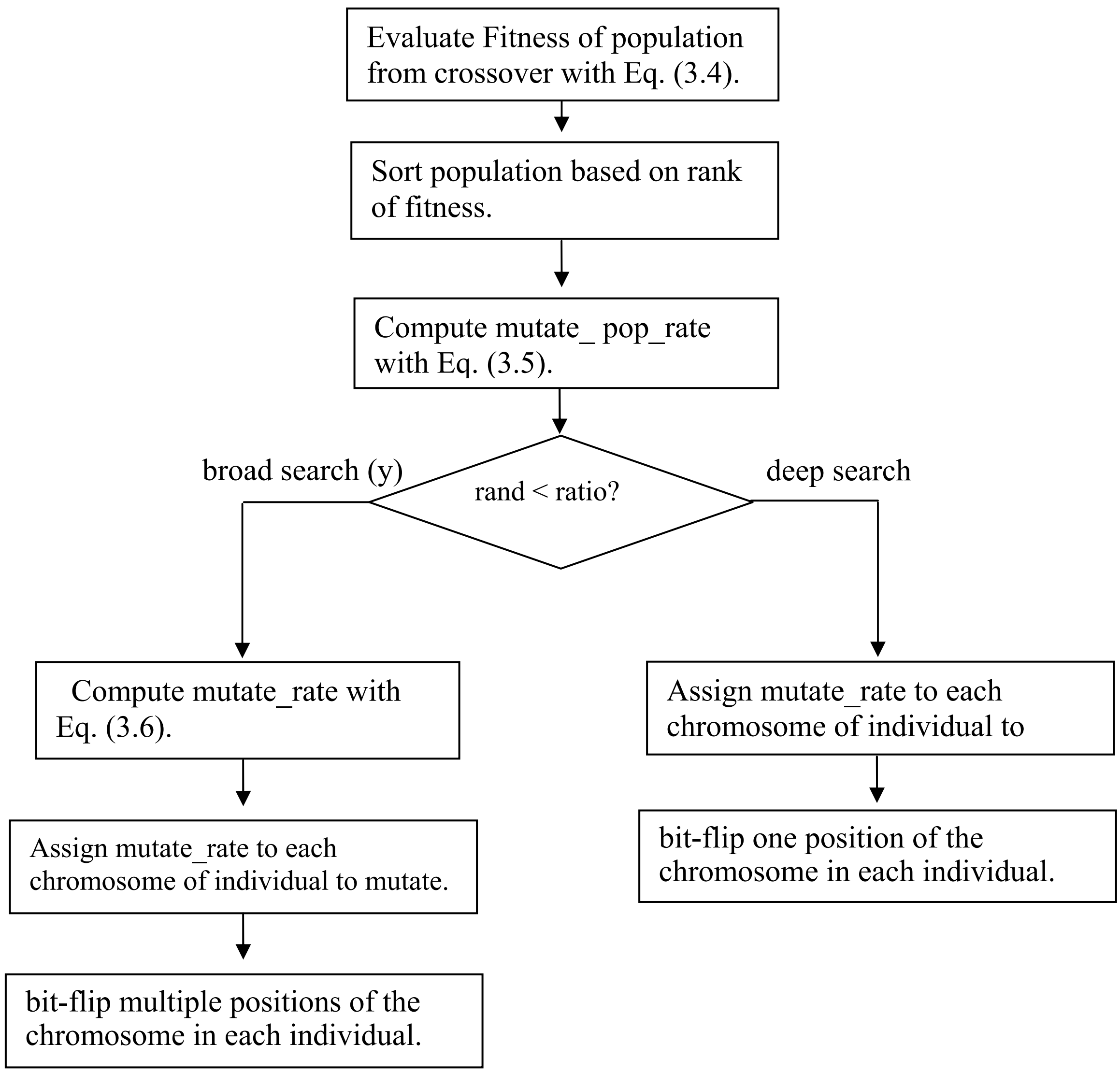

Figure 3.4. The procedure of broad search and deep search mutation operator

$$HDR = \frac{2}{|P||P-1||L|} \sum_{j=1}^{|P-1|} \sum_{k=j+1}^{|P|} \sum_{i=1}^{|L|} (s_{ji} - s_{ki})^2 \tag{3.7}$$

Where, |L| is the length of the bit string of an individual, in our case, it's equal to 20*9 = 180, |*P*| is the population size. sji and ski stands for the bit value, which is either 0 or 1, of position i in different individual j and k, respectively. If the bit value of the same position in different individuals is the same, $(s_{ji} - s_{ki})^2$ will be 0, otherwise it will be a 1. For a population of |P|, there are a total number of $\frac{|P||P-1|}{2}$ choices of asymmetry comparisons. Hence, HDR here means the average proportion of positions of different value in any two individuals of different order in the population of the same generation. Note that the value of HDR is in the range of [0, 0.5] and fits quite well with ordinary Min MAPE*10 in the graph used in this study.

For OGA, after generations of genetic evolution, homogeneity of genes will be increasing, and the corresponding HDR will be decreasing.

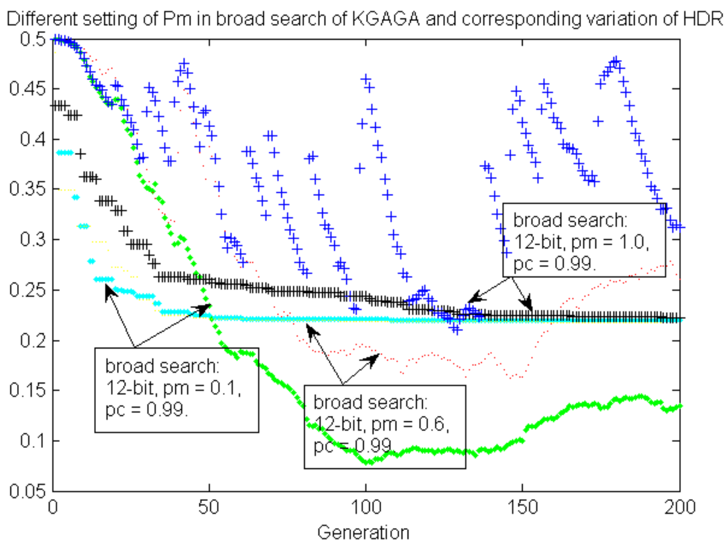

Figure 3.5. The influence of mutation rate on Min MAPE and *HDR* of KGAGA in broad search

3.4.2 P_m , P_c and number of bits to be mutated in broad search and HDR

Figure 3.5 shows the impact of variation of mutation rate (Pm) on HDR and min MAPE of KGAGA. There are three different Pm in broad search, namely 0.1 (big dots), 0.6 (small dots), and 1.0 ("+" marks) respectively. In general, the higher Pm is, the higher HDR jumps from the bottom (determined by pitfall) detected by DEMA. For instance, there is a large jump of HDR from each pitfall detected by DEMA as can be seen in "+" marks in Figure 3.5, even though the step size of each successive jumps decayed a bit. On the other hand, the step size of jump for setting Pm = 0.1 from the pitfall is very small as can be seen in green big dots in the same figure.

To show the influence of crossover rate (P_c) of broad search upon HDR, a figure was drawn at Figure 3.6, in which 3 different settings of P_c were set up. Among which, P_c = 0.1 was represented by big dots, P_c = 0.5 was represented by small dots, while P_c = 0.5 was represented by "+" mark. Each of the more straight curves from left to right (the higher two are overlapped) between 0.2 and 0.25 denotes Min MAPE * 10 of OGA. Many lines of dots or "+" marks go from top downwards a little to the right are curves of HDR. It demonstrates clearly that, in general, the higher the P_c is in broad search, the more local pitfalls will be found due to a convergence process of a higher pressure of selection triggered, resulting in more convergence processes, check Figure 3.6.

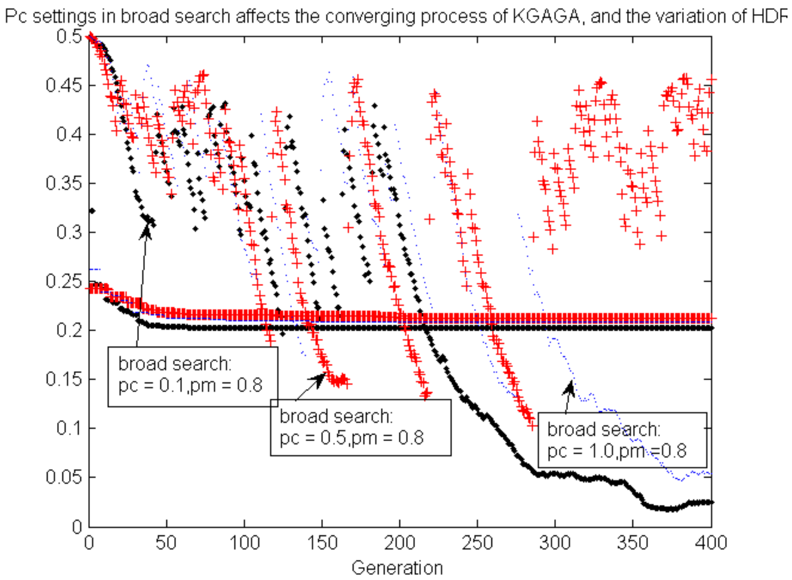

Figure 3.6. The influence of crossover rate in broad search on Min MAPE and *HDR* of KGAGA

To demonstrate the influence of number of bits to be mutated in broad search on HDR, three different settings of number of bits are set up in Figure 3.7, namely 1 bit (denoted by big dots), 5 bits (denoted by small dots), and 10 bits (denoted by circles) respectively. Figure 3.7 clearly shows that as the number of bits to be mutated from 1 bit to 5 bits and to 10 bits, the magnitude of variation of HDR from the bottom of a convergence (pitfall) to the top of another convergence increases also. This is quite similar to what happens to HDR when Pm increases in Figure 3.5.

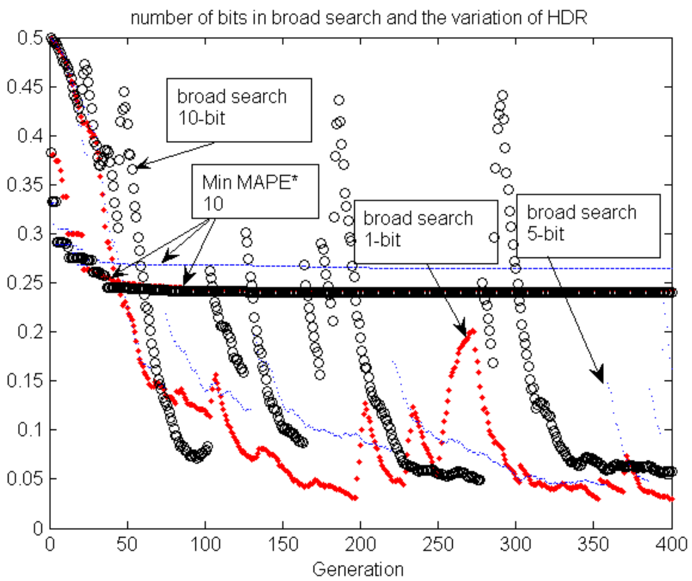

Figure 3.7. The impact of number of bits to be mutated in broad search on *HDR* and Min MAPE of KGAGA.

3.4.3 P_m and P_c in normal search and HDR

Since the proportion of broad search in the whole search process is only a small one, after broad search or deep search, normal search plays a very important role in the search of better solutions and convergence toward optimal or near optimal solution(s). Both mutation rate (P_m) and crossover rate (Pc) is critical in normal search.

Figure 3.8 shows three different settings of mutation rate in normal search of KGAGA for parameter estimation of a typical item. Although their min MAPE are all closely overlapped in the figure, the setting of P_m = 0.1 shows a better search ability, hence fewer bottlenecks (*MAFI(l=7)* = 0) are encountered. The search ability of setting of P_m = 0.01 and 0.001 is not as good as that of the setting of P_m = 0.1, there are much more local pitfalls arising in Figure 3.8. Note that, however, more convergence process doesn't guarantee better solution derived, because some convergence processes are not productive, please check Figure 3.10 below.

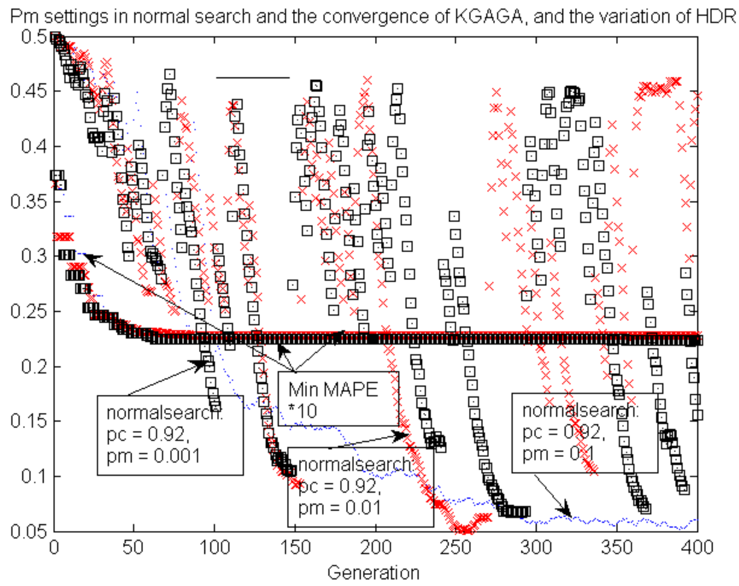

Figure 3.8. The influence of mutation rate in normal search on Min MAPE and *HDR* of KGAGA.

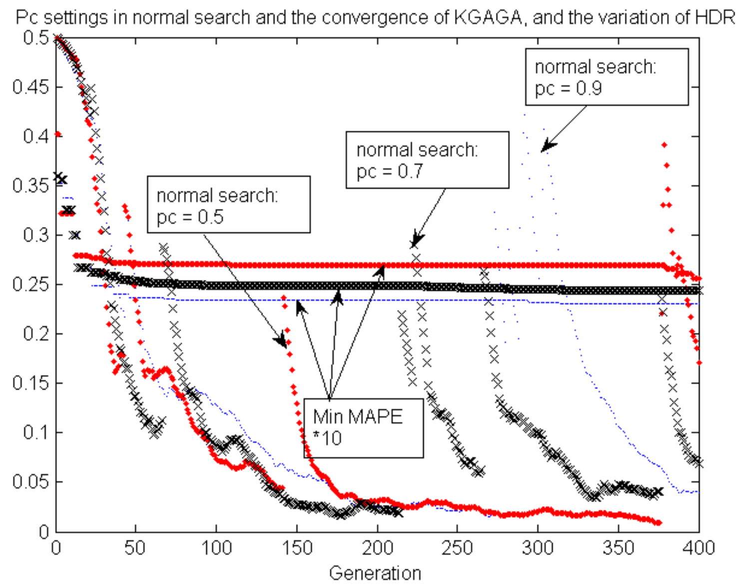

Figure 3.9. The influence of crossover rate in normal search on Min MAPE and *HDR* of KGAGA.

In Figure 3.9, the effect of three different settings of crossover rate in normal search upon Min MAPE *10 and HDR of KGAGA is demonstrated. All Pm are set at 0.01 in normal search. In general, the higher P_c is in normal search, the higher the pressure of selection, making a faster convergence, thus creating more bottlenecks and more convergence processes in the search process. For instance, in Figure 3.9, there are 4 convergence processes for the setting of P_c = 0.5, there are 6 convergence processes for both setting of P_c = 0.7 and P_c=0.9.

3.4.4 DEMA and HDR

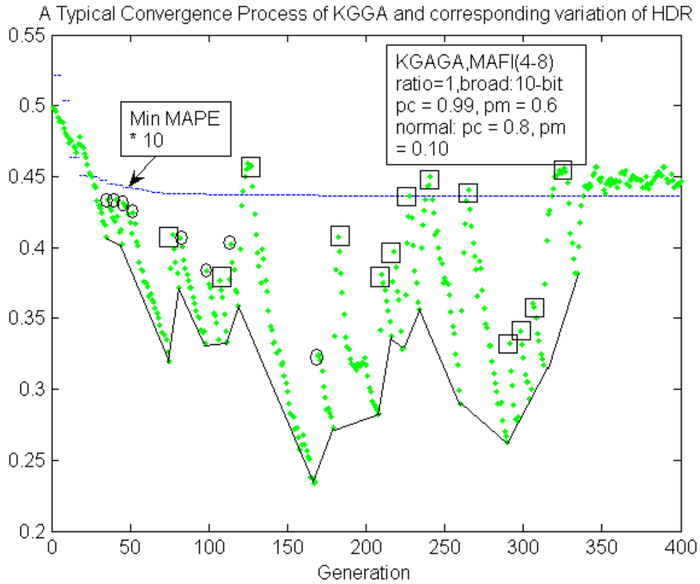

Figure 3.10. The pitfall detected by DEMA and the dramatic change of HDR thereafter.

To demonstrate about what value of HDR, DEMA detects a local pitfall in general and the variation of HDR after action taken by DEMA through a combination of a broad search mutation operator and a deep search mutation operator, Figure 3.10 and 3.11 were drawn. The connected line in the bottom of HDR in both figures demonstrates pitfall fronts detected by DEMA, while each of three circles represents start of an effective (contributing significant improvement in fitness) convergence process after actions taken by DEMA. Each of square boxes represents start of an un-effective (contributing no significant improvement in fitness) convergence process. The dramatic change of HDR from pitfall front to a circle or square in this figure is caused by a 10-bit broad search mutation operator with a Pm of 0.6. Note that in Figure 3.10 most of the re-convergences in the first quarter of search process are effective, while most of them are ineffective after 150 generations. Compared to its counterpart in Figure 3.10, DEMA in Figure 3.11 seems to be more effective in improving fitness of the best solution found so far, check that there are 7 circles and only one square in Figure 3.11.

To show the effectiveness of DEMA in terms of substantial content change of parameters derived in real number, Figure 3.12 is displayed, in which, partial outputs of KGAGA can be seen. In generation 25, a local pitfall is detected by DEMA, DEMA responds with a 12-bit mutation, causing parameter such as μ_2 and μ_3 to have a significant change in value. In generation 31, another pitfall was detected again, after a 12-bit mutation, causing a small change in the value of μ_2 and a significant change of μ_3, which, however, leads to a dramatic change of the value of price elasticity and μ_1 in generation 33. These substantial changes of parameters' value after multi-bit mutation evidenced the escape from local pitfall via DEMA in the course of KGAGA search process.

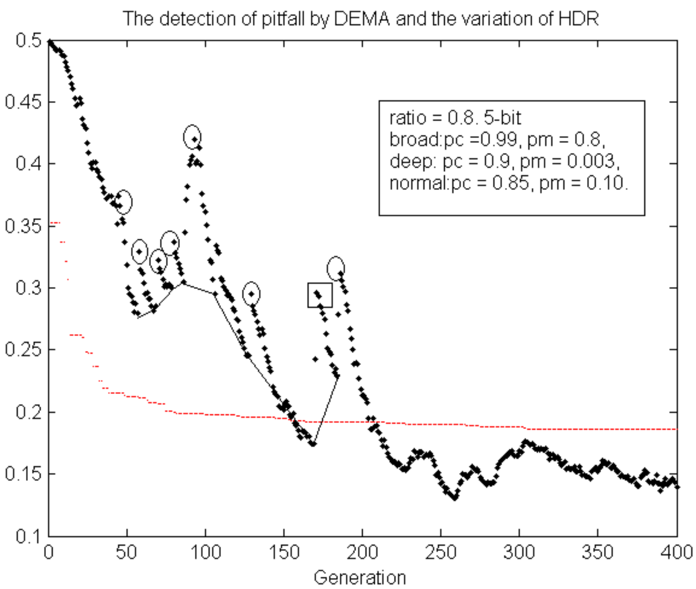

Figure 3.11. The pitfall detected by DEMA and the dramatic change of HDR thereafter.

```
Gen = 23, best MAPE = 0.0265946, PE = -0.282097, μ1 = 1.421193, μ2 = 2.688431, μ3= 1.657177,
Gen = 24, best MAPE = 0.0265946, PE = -0.282097, μ1 = 1.421193, μ2 = 2.688431, μ3= 1.657177,
Gen = 25, best MAPE = 0.0265946, PE = -0.282097, μ1 = 1.421193, μ2 = 2.688431, μ3= 1.657177,
***MUTATING type 3dc12
Gen = 26, best MAPE = 0.0262957, PE = -0.282173, μ1 = 1.450343, μ2 = 2.192608, μ3= 1.446905,
Gen = 27, best MAPE = 0.026006, PE = -0.280876, μ1 = 1.421193, μ2 = 2.688431, μ3= 1.657177,
Gen = 28, best MAPE = 0.0259558, PE = -0.282097, μ1 = 1.421193, μ2 = 2.687899, μ3= 1.044078,
Gen = 29, best MAPE = 0.0259558, PE = -0.282097, μ1 = 1.421193, μ2 = 2.687899, μ3= 1.044078,
Gen = 30, best MAPE = 0.0259558, PE = -0.282097, μ1 = 1.421193, μ2 = 2.687899, μ3= 1.044078,
Gen = 31, best MAPE = 0.0259558, PE = -0.282097, μ1 = 1.421193, μ2 = 2.687899, μ3= 1.044078,
***MUTATING type 3dc12
Gen = 32, best MAPE = 0.0255532, PE = -0.282097, μ1 = 1.421193, μ2 = 2.688431, μ3= 1.446905,
Gen = 33, best MAPE = 0.0245139, PE = -0.924188, μ1 = 1.171254, μ2 = 2.693097, μ3= 1.435967,
Gen = 34, best MAPE = 0.0243611, PE = -0.924188, μ1 = 1.171254, μ2 = 2.693097, μ3= 1.435967,
```

Figure 3.12. Detailed changes of values of parameters after multi-bit mutation triggered by DEMA.

3.4.5 A comparison of KGAGA and OGA

To give a basic idea about the fundamental difference of corresponding HDR in converging process of KGAGA and its counterpart of OGA, Figure 3.13 is drawn. Note that there are about 6 re-convergence processes in the first 100 generations of KGAGA in Figure 3.13, this is because we found that

DEMA is most effective in bout the first quarter of the whole search process, therefore in that period we usually set *l* in *MAFI(l)* to be a bit smaller than usual such as 4 to take full advantage of DEMA. Whereas corresponding HDR of converging process lingers in the range of 0.15 to 0.25 most of the time after about 50 generations of run, min MAPE.* 10 of OGA also looks almost constant thereafter. While Min MAPE of KGAGA keeps improving in the course of the search due to functioning of DEMA.

There is also a comparison of converging process between KGAGA and OGA, the result can be portrayed in Figure 14, which shows the capability to keep escaping from many local optimal solutions in the search process of KGAGA (denoted by large dots) with a significantly and consistently better performance within a much narrower band than that of OGA (denoted by small dots) which usually gets stuck quite early at local optimum with unstable performance result in a much wider band, check Figure 3.14.

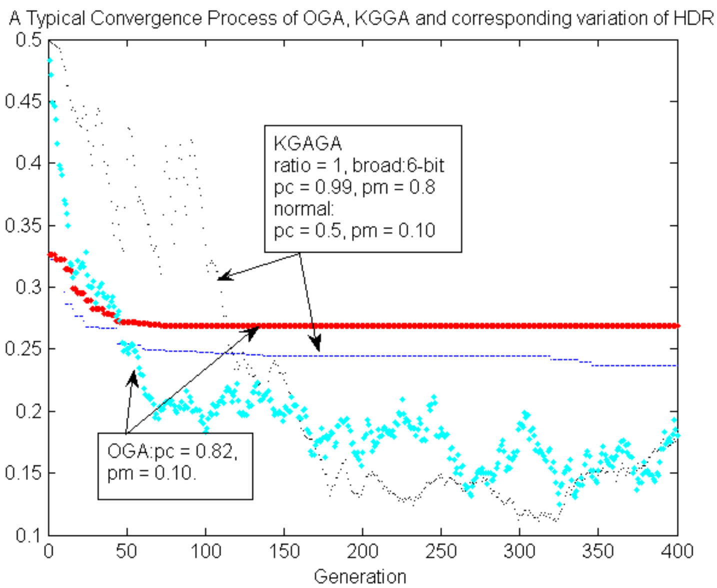

Figure 3.13. Typical converging processes of KGAGA and OGA, and their corresponding variations of HDR

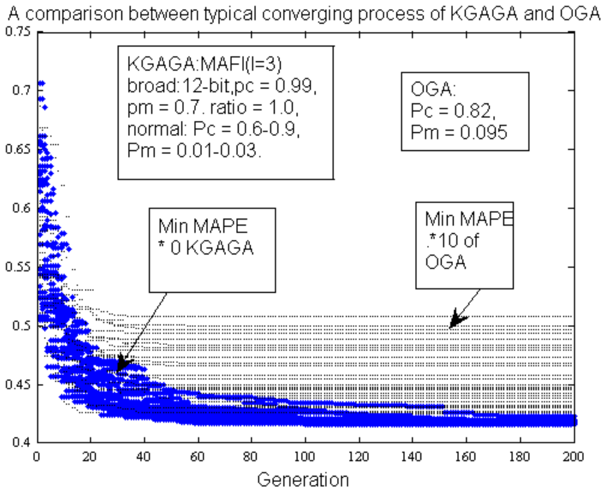

Figure 3.14. A comparison of the converging process between KGAGA and OGA for running 30 times of parameter estimation based on historical data of a typical item.

Table 3.1. Statistics of Min MAPE from running 30 times in parameter estimation based on sales data of an item with KGAGA and OGA respectively.

Min. MAPE	KGAGA	OGA
Avg	0.041782	0.045654
Max	0.042260	0.050806
Min	0.041600	0.042522
S. d.	0.000132	0.002569
IMP	8.48%	--

Table 3.2. Statistics of Min MAPE from running 15 times in parameter estimation based on sales data of an item with KGAGA and OGA respectively.

Min MAPE	KGAGA	OGA
Avg	0.0186684	0.0211797
Max	0.0191879	0.0250450
Min	0.0183177	0.0183457
s. d.	0.0002294	0.0024226
IMP	11.9%	--

The statistics of KGAGA and OGA in Table 3.1 and Table 3.2 further prove that the performance of KGAGA is about 8.48% and 11.9% better than that of OGA, respectively, in terms of average Min MAPE. Besides, The Maximum of Min MAPE of OGA in both table are much larger than that of KGAGA, while the Minimum of Min MAPE of OGA is slightly higher than that of KGAGA, in Table 3.1 it even is higher than that of Maximum for KGAGA, signifies an inferior and unstable performance of OGA. Table 3.2 provides statistics of results from parameter estimation between KGAGA and OGA on a typical item in CPG industry. In which, KGAGA also outperforms OGA in each aspect.

3.5 Model Checking

Regression diagnostics focused on normality and independence is performed to see if critical assumptions of linear regression are violated, based on Eq. (3.2). If these assumptions are severely violated, particularly if collinearity arises among predictor variables, variance of parameters may be a serious issue in model fitting or even in model specification.

Normality test is conducted through One-Sample Kolmogorov-Smirnov test (Lilliefors, 1967), and Q-Q plot (Berilant et al., 2005). Independence test in this thesis consists of two parts, namely, collinearity test and autocorrelation test. The former is performed via variance inflation factor (VIF), whereas the latter is performed via Durbin-Watson (D-W) test (Savin and White, 1977; Draper and Smith, p179-197, 1998).

Coenders and Saez (2000) addressed that VIF can be obtained as a measure of the increment of the sampling variance of the estimated regression coefficient of predictable variables due to collinearity. Vif_j can be computed as the jth diagonal value of the inverse of the R correlation matrix among the regressors or alternatively as $1/Tol_j$. Values of Vif_j lower than 10 or values of Tol_j larger than 0.1 are usually considered to be acceptable. Rook et al. (1990) also stated that *VIF* in excess of 10 indicates severe collinearity which leads to unstable estimation of the associated least squares regression coefficients.

VIF is one of the most popular measures used to detect collinearity in the literature (Belsley et al., 1980; Belsley, 1982; Stine, 1995), which can be derived via regression of one predictor variable to all other predictors and can be formulated as

$$VIF_{\mathrm{j}} = 1 \;/\; (1-R^2_{\mathrm{j}}). \qquad j = n + 2. \tag{3.8}$$

Where, n denotes the number of types of non-price promotion mixes specified in Eq. (3.1). R_{2j} is the coefficient of determination from regression of the *j-th* predictor variables on the other predictor variables. As mentioned in Theil (1971), estimated effect parameters can be directly proportional to VIF_{j} as the following equation:

$$s^2(\hat{b}_j) = VIF_j(\sigma^2/(T-1)V^2{}_j) \tag{3.9}$$

Where, $\hat{b}_j$ denotes the *j-th* effect parameters in Eq. (3.2). $s^2(\hat{b}_j)$,σ^2, and $V^2{}_j$ is the variance of $\hat{b}_j$.and variance of regression errors, as well as the variance of the *j-th* predictor variable, respectively. *T* denotes the number of periods in the training period, and may be perceived as sample size.

D-W test puts focus on testing whether there is any autocorrelation among the following series of regression error terms in Eq.(3.2) : $\varepsilon_{it}, \varepsilon_{it-1}, \ldots, \varepsilon_{i1}$. The statistic can be formulated as

$$d_i = \sum_{t=2}^{T}(\varepsilon_{it} - \varepsilon_{it-1})^2 / \sum_{t=1}^{T}\varepsilon^2{}_{it} \tag{3.10}$$

Where, *i* denotes an item *i*. In general, as the value of *d* increases to approximate 2, serial correlation is not a problem. If $d_i < dL$, the null hypothesis is rejected, and autocorrelation is serious. If $d_i > dU$, the null hypothesis is not rejected. However, if $dL <= d_i <= dU$, the test result is inclusive.

3.6 The Re-composition of Variable Coefficients Estimated

As the cycle length of CPG industry is one year, about 52 weeks long, let *t'* = *t* + 52, denoting the corresponding week to be forecasted in a new year. Modified from Eq. (3.1), a modified naïve sales forecasting method considering cycle length to forecast unit sales of item *i* of period *t'* in a new year, based on sales data of the same week t in the referenced year, would be

$$\hat{S}_{it'} = \eta_i \pi_{it} \left(P_{it'} / \widehat{P}_i\right)^{\theta_{i'}} \prod_{l=1}^{n} \mu_i^{D_{lit'}} \beta_i^{E_{it}}, t' = t + 52, \forall t' \in Z. \tag{3.11}$$

Where, η_i denotes the geometric average of normal sales volume of a certain item *i* across *N* periods of time in the referenced period. While π_{it} denotes the seasonal index for item *i* in period *t*. So far, all the parameters in Eq. (3.11) are already derived. Let $e_{1it'}$ denotes the price effect multiplier of item *i* in forecasting period *t'* and $e_{2it'}$ denotes the effect multiplier of a non-price promotion mix. While $e_{3it'}$ denotes the decay effect multiplier of item *i* in forecasting period *t'*. In each group of indicator parameters at most one condition will arise in each period. We get

$$\hat{S}_{it'} = \eta_i \pi_{it} e_{1it'} e_{2it'} e_{3it'} \qquad , t' = t + 52, \forall t' \in Z. \tag{3.12}$$

In its re-composed form, Eq. (3.12) can be used in responding to expected promotional campaigns in the forecasting horizon, as specified in the promotion proposals, to perform multi-step out-of-sample forecasting without any adjustment in the following empirical study.

3.7 The Measurement of Forecasting Errors

Accuracy is one of the most important criteria to compare and select forecasting methods. To draw conclusions about forecasting accuracy usually need to compare across a wide variety of time series, however, it is usually hard to obtain a large number of series. In the literature, some of the most frequently used measures to assess forecasting accuracy are RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), MdAPE (Median Absolute Percentage Error), GMRAE (Geometric Mean Relative Absolute Error), and MdRAE (Median Relative Absolute Error). In general, each measure has its own strength and weakness. Armstrong and Collopy (1992) conducted an empirical study and concluded with the following Table 3.3.

These error measures are defined below:

Let $e_t = S_t - \hat{S}_t$, denoting the forecast error.

$r_t = e_t / e*_t$, denoting the relative error where $e*_t$ is the forecast error derived from the benchmark method like NAÏVE or other alternatives.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} e_t^{\,2}}{n}} \tag{3.13}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{S_t}\right| \tag{3.14}$$

$$MdAPE = median\left(\left|\frac{e_t}{S_t}\right|\right) \tag{3.15}$$

$$GMRAE = \sqrt[n]{\prod_{t=1}^{n} |r_t|} \tag{3.16}$$

$$MdRAE = median(|r_t|) \tag{3.17}$$

Table 3.3. Ratings of the forecasting error measures*

Error measure	Reliability	Construct validity	Outlier protection	Sensitivity	Relationship to decisions
RMSE	Poor	Fair	Poor	Good	Good
Percent Better	Good	Fair	Good	Poor	Poor
MAPE	Fair	Good	Poor	Good	Fair
MdAPE	Fair	Good	Good	Poor	Fair
GMRAE	Fair	Good	Fair	Good	Poor
MdRAE	Fair	Good	Good	Poor	Poor

* Armstrong and Collopy (1992)

Most textbooks recommend the use of MAPE, see Hyndman and Koehler (2005). Actually, MAPE is the most popular error measure used in practice. In this thesis, MAPE is used for comparisons of multi-step out-of-sample weekly sales forecasting accuracy based on re-combination of parameters, in Eq. (3.12), derived from LS estimator and KGAGA estimator in model fitting.

4. Empirical Study—Part 1

4.1 Background

This thesis has a focus on the forecast of weekly unit sales of several series of products, manufactured by a branded CPG Company A, under retailer B. Company A is a leading manufacturer specialized in dehumidifier and deodorizer products in Taiwan. While retailer B is an international outlet of DIY products.

A sales data set of 30 items in 2007, aggregated from retailer B's outlets, coupled with price promotion, non-price promotion data, as well as promotion proposals of the first 3 months in 2008, which were set up more than three months earlier in 2007, are used to conduct our empirical study. The details of price rate and type of non-price promotion mix of these items are displayed in Table 6.1. Based on these promotion proposals, forecasters are able to know in advance detailed changes of promotion activities in the forecasting horizon such as the information of price rate, the type of promotion mix, and the starting and the ending of a promotion session.

KGAGA was recommended to assess effect parameters of the model. The underlying equation is Eq. (3.2). The fitness function concerned is Eq. (3.4). Each effect parameter is defined with a constraint in KGAGA implemented in Matlab 7.1. For example, the price elasticity coefficient is set to be in the range of [-9, 2], while effect parameters of non-price promotion mixes are set to be within [1, 5]. However, the coefficients of predictor variables in OLS regression are estimated without any constraint in enter mode, which is most similar to KGAGA in variable selection, in SPSS in that it doesn't remove any insignificant parameter in general. The type number of non-price promotion mixes n in Eq. (3.2) is set to be 6 to reflect the business reality. The dummy variable of decay effect after week 4 of a promotion session is set to be 1 if average sales volume of the first 4 weeks in promotion is more than 10% above that of the remained weeks in promotion, otherwise 0. Overall, parameters estimated are used to be recomposed like Eq. (3.12) according to the price rate and type of non-price promotion mix in the promotion proposals.

4.2 Design of the empirical study

To take both the busy season and off season into account to have a proper assessment of the performance of both methods, the forecasting horizon is designed to be composed of two periods of equal length, in which the first period includes the first 6 weeks of 2008 which can be denoted as busy season, while the second one starts from the 11th week and ends at the16th week of 2008, which can be denoted as off season.

To properly evaluate the performance of parameter estimation via KGAGA and OLS as well as that of multi-step out-of-sample forecasting based on parameters derived from KGAGA (denoted as KGAGA model) and OLS in enter mode (denoted as OLS model), respectively, particularly the consistency of performance, we conduct model fitting and model checking consecutively with KGAGA, then with OLS in enter mode, all based on the dataset of the whole year of 2007 as the first training period, training dataset in this period can be denoted as sample 1. Besides, the dataset of year 2007 and the first 10 weeks of 2008 is the second training period, training dataset in this period can be denoted as sample 2.

Thereafter, parameters derived from either estimator based on either sample 1 or sample 2 are recomposed without adjustment to forecast weekly unit sales of items concerned in the forecasting horizon.

To further investigate the relative forecasting performance of our regression model estimated with KGAGA and other popular forecasting alternatives like ARIMA, Exponential smoothing (ES), and Naïve model considering the cycle length of 52 weeks. In this study, we also make a comparison between MAPE of KGAGA model and that of ARIMA, ES, and Naïve in both sample 1 and sample 2 respectively in order to shed some light on their relative performance in both busy and off seasons.

A brief description of these forecasting alternatives can be stated below:

ARIMA

ARIMA stands for Autoregressive integrated moving average, it's a generalization of ARMA (autoregressive moving average), and it is fitted to time series data for analytical and forecasting purposes. A differencing step can be used to remove the non-stationarity of data when it's needed.

The model is generally referred to as ARIMA (p,d,q), where p, d, and q are integers greater than or equal to 0 and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{4.1}$$

Where, X_t represents a time series of data, in which t stands for index of time interval, L is the lag operator, the α_i are parameters of the autoregressive part of the model. The θ_i are parameters of the moving average part and the ε_t are error terms which are assumed to be conformed with normal distribution with zero mean in general.

Assuming the polynomial $\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)$ has a unitary root of multiplicity d, then it can be rewritten as

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)=\left(1+\sum_{i=1}^{p-d}\phi_i L^i\right)(1-L)^d \tag{4.2}$$

This polynomial factorization property can be expressed by an ARIMA (p,q,d) and is given by:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{4.3}$$

ARIMA can be viewed as a cascade of two models. The first is non-stationary:

$$Y_t=(1-L)^d X_t \tag{4.4}$$

While the second is wide-sense stationary:

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)Y_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t \tag{4.5}$$

4.3 Empirical Results

In this section, the results of model fitting, and model checking of the regression model, as well as forecasting results of KGAGA model, OLS model, and ARIMA and ES are delineated in detail and analyzed in a comparative fashion.

4.3.1 The results of model fitting

Most parameters derived from KGAGA are consistent with our expectations, for example, the parameters of non-price promotion activities are all at least equal to 1.0, the parameters of price discount (price elasticity) are at most to be zero, most of them are negative which are reasonable.

Overall, effect parameters estimated by OLS also are roughly consistent with our expectations, except that some magnitudes of these parameters are much smaller than expectation, they are marked with bold type, for example, μ_2 , μ_4 and μ_6 in item 1, μ_1, μ_2 , and μ_6 of item 24 in Table 4.3, some price elasticity are more than 1 significantly, like that of item 8 and item 26 in Table 3.6. others such as μ_2 of item 9, μ_1, μ_2 , and μ_6 of item 14, μ_6 of item 21, and item 24 as well as price elasticity (θ) of item 25 and 26 in Table 4.4 are deserved to mention.

In contrast, parameters estimated by KGAGA which takes into considerations of contextual knowledge gained from the literature and life experience and realize it in terms of constraints so as to be more true to the reality, even though some numbers derived are still at odds with our expectations, however, they are not as severe as those estimated by OLS as stated above, check Table 4.1-4.2.

Parameters of non-price promotion mixes being less than 1 imply negative effect of non-price promotion, similarly, a price elasticity greater than 1 implies higher price induces more customers to buy, both sounds in conflict with our experience.

These abnormal phenomena arose in parameters measured with OLS which is usually tormented with multi-collinearity and auto-correlation issues, warrant a further investigation.

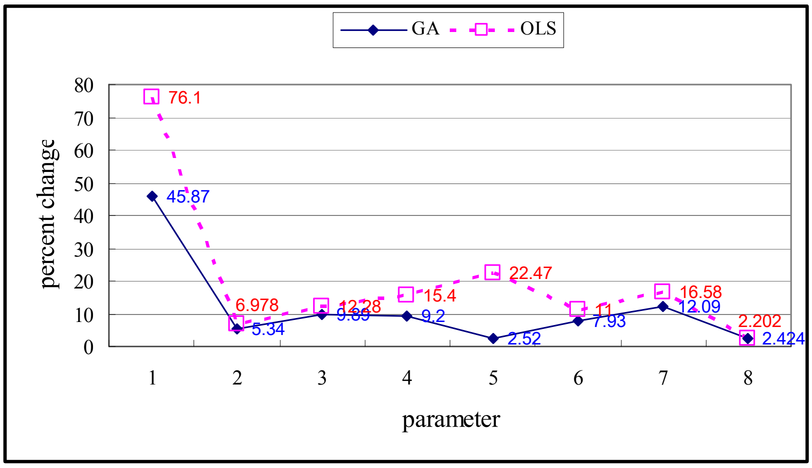

Figure 4.1. A comparison of percent change of parameters derived from KGAGA and OLS on different size of samples

To better understand the stability of parameters measured by KGAGA and OLS, we compare the stability (measured with percent change) of parameters derived from these estimators on different dataset of different sample size, percent change (see Eq. (4.6) in pp. 58) is the average absolute difference of two numbers derived from dataset of two different sample sizes divided by the average of these two numbers of the same parameter across items.

Table 4.1. Parameters estimated via KGAGA on sample 1.

Item (i)	**MAPE$_i$**	**Price elas.θ_i**	**base sales λ_i**	**P-mix $\mu_{1i(t=1:T)}$**	**P-mix $\mu_{2i(t=1:T)}$**	**P-mix $\mu_{3i(t=1:T)}$**	**P-mix $\mu_{4i(t=1:T)}$**	**P-mix $\mu_{5i(t=1:T)}$**	**P-mix $\mu_{6i(t=1:T)}$**	decay rate β_i
1	0.0515	-2.270	53.408	1.273	1.380	--	1.500	--	1.727	0.734
2	0.0428	-7.557	39.002	1.123	--	1.625	--	3.091	--	0.926
3	0.0167	-0.429	186.23	1.004	--	--	1.420	**0.886**	--	0.952
4	0.0235	-8.731	155.87	1.121	--	1.533	--	1.086	--	1.000
5	0.0279	0.000	107.00	1.000	--	1.598	--	1.000	--	1.000
6	0.0279	0.000	106.89	1.000	3.834	1.600	--	1.000	--	1.000
7	0.0365	-2.191	79.476	1.0695	--	1.941	--	1.460	--	1.000
8	0.0319	0.000	99.487	1.800	2.060	1.025	--	2.222	--	0.742
9	0.0279	0.000	378.25	1.000	--	1.598	--	1.000	--	1.000
10	0.0183	-0.293	370.16	1.463	--	--	2.313	1.268	1.423	0.862
11	0.0246	-2.725	105.42	1.238	1.339	1.318	--	1.384	2.254	0.844
12	0.0263	-1.280	106.22	1.429	1.826	2.005	--	--	1.201	0.813
13	0.0207	-0.975	232.35	1.464	2.265	2.058	--	--	1.312	0.813
14	0.0253	0.000	264.60	1.019	**0.877**	1.686	--	--	1.191	0.882
15	0.02	-0.123	73.232	1.099	1.195	--	1.409	--	--	1.000
16	0.0333	-0.100	135.10	1.707	--	--	1.756	--	1.654	0.676
17	0.0351	-2.753	74.019	1.208	1.558	1.206		2.371	--	1.000
18	0.0245	-1.775	142.53	1.852	--	--	1.000	1.698	--	1.000
19	0.0338	-1.016	163.38	1.000	1.530	1.806	--	1.090	--	1.000
20	0.031	-0.529	126.14	1.766	2.989	1.000	--	--	3.229	1.000
21	0.0174	0.000	219.38	**0.800**	1.188	1.522	--	--	1.272	0.875
22	0.0312	-1.699	92.168	1.000	1.226	1.736		3.016	--	1.000
23	0.0416	-0.221	63.864	1.292	2.955	--	1.730	2.144	--	0.969
24	0.0366	-1.522	91.678	0.931	1.035	1.818	--	--	1.020	0.870
25	0.0282	-0.090	155.97	--	1.316	1.692	--	--	1.239	0.797
26	0.0448	0.000	74.300	1.000	1.359	3.149	--	--	--	1.000
27	0.0456	-0.076	68.420	1.000	1.505	3.075	--		--	1.000
28	0.0219	-0.110	89.030	1.620	1.582	1.382	--	1.566	1.187	0.549
29	0.0219	-3.080	107.99	1.470	1.324	1.537	--	1.164	1.000	0.740
30	0.0185	-0.825	130.42	1.306	1.548	1.235	--	1.395	1.563	0.781
mean	0.030	-1.597	127.358	1.243	1.709	1.702	1.590	1.602	1.519	0.894

Table 4.2. Parameters estimated via KGAGA on sample 2.

item (i)	**MAPE$_i$**	**Price elas.θ_i**	**base sales λ_i**	**P-mix $\mu_{1i(t=1:T)}$**	**P-mix $\mu_{2i(t=1:T)}$**	**P-mix $\mu_{3i(t=1:T)}$**	**P-mix $\mu_{4i(t=1:T)}$**	**P-mix $\mu_{5i(t=1:T)}$**	**P-mix $\mu_{6i(t=1:T)}$**	decay rate β_i
1	0.0521	-4.185	58.410	0.937	0.952	--	1.330	--	1.190	0.766
2	0.0407	-6.990	41.158	1.141	--	1.922	--	2.942	--	0.930
3	0.0174	-0.361	183.83	1.047	--	--	1.508	0.908	--	0.953
4	0.0245	-6.833	167.99	1.134	--	1.580	--	1.100	--	1.000
5	0.0279	0.000	107.00	1.000	--	1.598	--	1.000	--	1.000
6	0.024	-0.600	102.20	1.000	1.253	1.558	--	1.178	--	1.000
7	0.0374	-0.376	85.125	1.000	--	2.569	--	1.504	--	1.000
8	0.0418	-0.559	106.68	1.678	1.716	1.625	--	1.852	--	0.774
9	0.0245	0.000	408.97	1.125	1.000	--	1.845	1.406	--	1.000
10	0.0193	-0.293	374.63	1.398	--	--	2.202	1.213	1.360	0.863
11	0.0250	-3.753	110.28	1.075	1.170	1.294	--	1.209	2.351	0.857
12	0.0340	-0.215	123.30	1.238	1.574	2.450	--	2.521	1.422	0.844
13	0.0223	-1.365	222.67	1.432	2.363	2.038	--	--	1.181	0.780
14	0.0245	-0.003	278.61	0.965	1.128	1.819	--	--	1.188	0.844
15	0.0207	-0.520	72.445	1.111	1.038	--	1.388	--	--	1.000
16	0.0365	-0.922	119.93	1.663	--	--	1.707	--	1.527	0.653
17	0.0369	-4.591	74.181	1.169	1.381	1.075	--	2.098	--	1.000
18	0.025	-1.761	143.06	1.845	--	--	1.000	1.694	--	1.000
19	0.0354	-2.324	136.28	1.100	1.834	1.674	--	1.090	--	1.000
20	0.0338	-0.543	132.14	1.677	2.852	1.000	--	--	3.032	1.000
21	0.0234	-1.014	217.52	0.828	1.084	1.496	--	--	**0.973**	0.877
22	0.033	-1.699	92.411	1.000	1.223	1.742	--	3.023	--	1.000
23	0.0416	-0.221	63.864	1.292	2.955	--	1.730	2.144	--	0.969
24	0.0404	-1.399	100.34	0.865	0.907	1.419	--	--	**0.957**	0.884
25	0.0301	-0.315	156.20	--	1.298	1.625	--	--	1.239	0.797
26	0.0477	0.000	77.43	1.000	1.279	2.805	--	--	--	1.000
27	0.0521	0.000	69.00	1.000	1.492	2.870	--	--	--	1.000
28	0.0256	-0.425	93.76	1.536	1.658	1.312	--	1.675	1.370	0.625
29	0.0219	-0.254	119.5	1.297	1.662	1.419	--	1.626	1.599	0.750
30	0.0234	-0.943	127.90	1.329	1.550	1.319	--	1.423	1.533	0.750
mean	0.031	-1.417	126.625	1.203	1.517	1.737	1.589	1.663	1.494	0.897

Note: T = 56 here.

Table 4.3. Parameters estimated via OLS on sample 1.

item (*i*)	**MAPE**$_i$	**Price elas.**θ_i	**base sales** λ_i	**P-mix** $\mu_{1i(t=1:T)}$	**P-mix** $\mu_{2i(t=1:T)}$	**P-mix** $\mu_{3i(t=1:T)}$	**P-mix** $\mu_{4i(t=1:T)}$	**P-mix** $\mu_{5i(t=1:T)}$	**P-mix** $\mu_{6i(t=1:T)}$	decay rate β_i
1	0.0531	-9.750	61.069	**0.447**	**0.401**	--	**0.712**	--	**0.524**	0.842
2	0.0445	-7.582	42.224	1.025	--	1.770	--	2.968	--	0.868
3	0.0182	-0.205	188.67	1.017	--	--	1.586	0.934	--	0.951
4	0.0241	-7.687	159.49	1.140	--	1.844	--	1.165	--	1.000
5	0.0255	0.386	115.01	0.872	--	1.654	--	0.882	--	1.000
6	0.0244	-0.559	106.91	0.847	1.195	1.694	--	1.150	--	1,000
7	0.0373	-3.009	81.37	0.99	--	1.763	--	1.373	--	1.000
8	0.0317	**1.242**	102.00	1.804	2.149	1.458	--	2.266	--	0.842
9	0.0234	0.000	132.42	0.882	**0.756**	--	1.530	1.204	--	1.000
10	0.019	0.132	374.65	1.531	--	--	2.517	1.320	1.411	0.840
11	0.0236	-2.638	129.28	1.034	1. 201	1.132	--	1.142	1.857	0.853
12	0.0276	-1.112	110.39	1.33	1.761	2.115	--	--	1.208	0.780
13	0.0216	-0.975	292.95	1.510	1.904	2.102	--	--	1.470	0.750
14	0.0262	0.000	299.47	0.829	1.000	1.687	--	--	1	0.922
15	0.0211	0.000	719.82	1.203	1.214	--	1.826	--	--	1.000
16	0.0315	-0.152	141.99	1.729	--	--	1.688	--	1.527	0.653
17	0.0377	-1.456	82.434	1.192	1.603	1.251	--	2.445	--	1.000
18	0.024	-2.350	140.89	1.865	--	--	0.851	1.618	--	1.000
19	0.0346	0.000	164.35	1.0253	1.552	2.387	--	1.278	--	1.000
20	0.032	-0.398	145.04	1.618	2.560	0.841	--	--	3.121	1.000
21	0.0177	0.334	236.99	**0.718**	1.110	1.655	--	--	1.300	0.926
22	0.0319	0.000	96.160	0.966	1.228	1.603	--	--	2.900	1.000
23	0.0476	-0.796	71.023	1.084	2.250	1.984	--	2.008	--	0.816
24	0.0342	-2.139	127.23	**0.628**	**0.724**	1.241	--	--	**0.568**	0.846
25	0.0291	-0.180	146.35	--	1.391	2.111	--	--	1.384	0.814
26	0.046	**1.190**	83.680	1.041	1.223	3.561	--	--	--	1.000
27	0.0477	**1.426**	68.100	1.132	1.525	4.154	--		--	1.000
28	0.0224	0.000	92.019	1.560	1.777	1.392	--	1.788	1.581	0.656
29	0.0211	-1.964	107.61	1.442	1.538	1.573	--	1.391	1.220	0.743
30	0.0195	-0.821	126.60	1.218	1.567	1.298	--	1.430	1.507	0.812
mean	0.030	-1.302	158.206	1.161	1.449	1.838	1.530	1.551	1.505	0.897

Note: T = 47 here.

Table 4.4. Parameters estimated via OLS on sample 2.

item (i)	$MAPE_i$	Price elas. θ_i	base sales λ_i	P-mix $\mu_{1i(t=1:T)}$	P-mix $\mu_{2i(t=1:T)}$	P-mix $\mu_{3i(t=1:T)}$	P-mix $\mu_{4i(t=1:T)}$	P-mix $\mu_{5i(t=1:T)}$	P-mix $\mu_{6i(t=1:T)}$	decay rate β_i
1	0.0516	-4.093	61.806	0.892	0.912	--	1.347	--	1.149	0.766
2	0.0431	-7.496	42.606	1.026	--	1.904	--	2.945	--	0.869
3	0.0185	-0.229	189.24	1.009	--	--	1.586	0.929	--	0.951
4	0.0253	-6.889	165.34	1.140	--	1.774	--	0.877	--	1.000
5	0.0255	0.386	115.70	0.867	--	1.606	--	1.149	--	1.000
6	0.0246	-0.377	108.64	0.834	1.189	1.636	--	1.178	--	1.000
7	0.0377	-1.281	85.970	0.945	--	2.396	--	1.411	--	1.000
8	0.0411	-0.686	115.12	1.474	1.550	1.380	--	1.800	--	0.773
9	0.0242	-1.757	459.44	0.965	**0.612**	--	1.239	1.318	--	1.000
10	0.0197	-0.315	380.32	1.400	--	--	2.206	1.232	1.283	0.827
11	0.0250	-2.638	121.27	1.102	1.195	1.328	--	1.218	1.982	0.853
12	0.0281	-1.112	110.38	1.330	1.761	2.361	--	2.314	1.208	0.780
13	0.0238	-0.475	284.58	1.756	1.962	2.275	--	--	1.311	0.741
14	0.0254	-3.454	308.90	**0.493**	**0.605**	--	--	--	**0.629**	0.922
15	0.0400	-0.787	694.37	1.247	1.412	--	0.931	--	--	1.000
16	0.0376	-0.407	122.98	1.684	--	--	1.876	--	1.623	0.676
17	0.0400	-5.041	82.434	1.182	1.269	0.951	--	1.935	--	1.000
18	0.0246	-2.480	141.03	1.863	--	--	0.841	1.587	--	1.000
19	0.0365	-2.326	154.78	1.089	1.649	1.610	--	0.862	--	1.000
20	0.0349	-0.398	140.05	1.675	2.651	0.966	--	--	3.22	1.000
21	0.0242	-2.750	225.88	0.822	0.931	1.219	--	--	**0.589**	0.860
22	0.0335	-1.469	96.159	0.966	1.700	1.228		2.895	--	1.000
23	0.0471	-0.738	71.880	1.078	2.255	--	2.246	2.002		0.815
24	0.0410	-2.139	103.44	**0.772**	0.890	1.214	--	--	**0.699**	0.846
25	0.0312	0.474	147.23	--	1.440	1.952	--	--	1.565	0.814
26	0.0465	**2.026**	87.182	1.064	1.168	3.203	--	--	--	1.000
27	0.0523	-0.076	70.598	1.071	1.471	3.040	--		--	1.000
28	0.0262	-0.818	93.500	1.536	1.603	1.370	--	1.553	1.300	0.742
29	0.0230	-0.607	112.28	1.413	1.660	1.478	--	1.632	1.654	0.740
30	0.0245	-0.821	122.98	1.254	1.614	1.508	--	1.473	1.551	0.812
mean	0.033	-1.626	167.203	1.171	1.432	1.733	1.534	1.595	1.412	0.893

Note: T = 56 here.

$$pc = \frac{|p_1 - p_2|}{\frac{p_1 + p_2}{2}} = \frac{2|p_1 - p_2|}{p_1 + p_2} \tag{4.6}$$

Where, pc represents the percent change, p1 and p2 represent a parameter estimated by OLS47 (OLS on sample 1) and OLS56 (OLS on sample 2), respectively.

It's obvious to see that KGAGA, in average, has a much more stable performance in parameter estimation than that of OLS from Figure 3.15. Because every point in this figure shows that percent change represented with blue line with diamonds caused by different sample size in parameters from KGAGA is smaller than its counterpart represented with dashed line with empty squares in OLS of Figure 4.1, except the point of parameter 8.

4.3.2 A comparative analysis of results of model checking via VIF and D-W test

The normality test, consisting of one-sample Kolmogorov-Smirnov test and Q-Q plot, in which, both KGAGA and OLS passed the test with data from either sample 1 or sample 2 without difficulty based on standardized error term ε_{it} in Eq. (3.2) and natural logarithm of predicted weekly unit sales denoted as $\ln \widetilde{S}_{it}$ in Eq. (3.2). However, in independence test, the measures of VIF and the D-W tests shown complex but interesting results in two training periods of different length via KGAGA and OLS and deserve some more investigation.

As shown in Figure 4.2, the blue dashed line with empty black square represent the maximal VIF of parameter of specific predictor variable regressed with other predictor variables of the model measured with OLS across 30 items on sample 1, whereas the purple line with triangles represents maximal VIF of parameter of predictor variable regressed with other predictor variables of the model measured with OLS across 30 items based on sample 2. Note that in Figure 3.8, there are three high points, the first one arises in item 1, the second one arises in item 21, the last one arises in item 25. In the first two cases, more observations seem to help alleviate the problem of multi-collinearity due to more information offered by these additional observations. While in case three, surprisingly, there is a rise of VIF owing to more observations, this may imply that more observations may in some case increase the propensity of some predictor variable to have multi-collinearity with other predictor variables.

Table 4.5. Maximal VIF across items in parameter estimation of OLS47 and OLS56

item	1	2	3	4	5	6	7	8	9	10
VIF-OLS47	114.96	3.026	1.916	2.291	1.95	2.419	1.49	4.517	2.668	5.56
VIF-OLS56	5.075	3.08	1.657	2.619	2.37	3.043	5.464	3.108	3.203	2.997

item	11	12	13	14	15	16	17	18	19	20
VIF-OLS47	4.813	9.234	7.246	1.199	1.076	2.908	5.119	6.15	1.723	2.011
VIF-OLS56	2.873	10.196	6.945	1.693	3.358	3.046	3.358	9.126	3.062	1.904

item	21	22	23	24	25	26	27	28	29	30
VIF-OLS47	90.233	1.432	1.664	8.113	1.9	3.499	3.364	2.273	2.273	4.50
VIF-OLS56	15.298	1.558	1.73	7.038	40.114	3.678	4.034	7.657	7.657	5.039

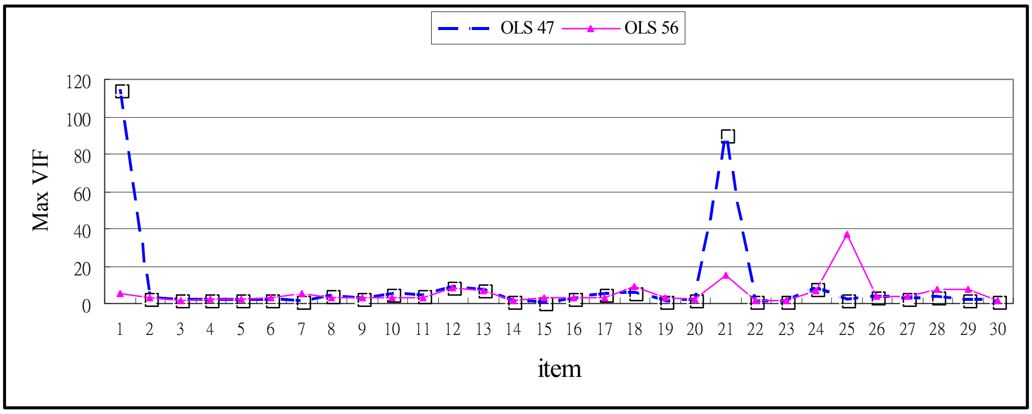

Figure 4.2. A comparison of VIF between OLS47 and OLS56

In general assessment of multi-collinearity, the cut point of VIF most often is set to be 10, see Rook et al. (1990), Belsley et al., (1980), note that in Table 4.5, there are two cases of OLS47 and three cases of OLS 56, all have max VIF greater than 10.

However, from a more conservative perspective, a cut point of 5 can be used to check if multi-collinearity is causing trouble, max VIF of item 1, 10, 12, 13, 17, 18, 21, and 24 in parameter estimation with OLS47 which is an estimator of OLS on sample 1 of 47 observations, while max VIF of item 1, 12, 13, 18, 21, 24, 25, 28, 29 and 30 in parameter estimation by OLS56 which is an estimator of OLS on sample 2 of 56 observations are all over 5.

To see if these large max VIFs in parameter estimation with OLS have any impact in the results of model fitting, check Table 4.3 to Table 4.4 as a reference. It’s actually difficult to set a clear cut point to identify the impact of multi-collinearity, but roughly speaking, larger max VIF in general did cause abnormal numbers derived in these two tables, see item 1, 8, 21, 24, 26, and 27 in Table 3.6. Also see item 18, 21, 24, 25, and 26 in Table 3.6 Some of them have price elasticity to be much greater than 0, and some of them have non-price promotion effect to be much less than 1.

Except a few items, generally speaking, the change of sample size from 47 to 56 didn't affect the performance of OLS's parameter estimation in enter mode too much, the number of large max VIFs remained about the same in both samples. Since these VIF measure collinearity among independent variables of the model estimated by OLS, which also reveal the same cillinearity of among the independent variables of the model estimated by KGAGA, it's not necessary to measure collinearity among independent variables of the model estimated by KGAGA.

In contrast, KGAGA shows a better and more consistent performance in model fitting. The problems caused by the presence of multi-collinearity among predictor variables of model based on smaller dataset didn't affect KGAGA too much in its parameter estimation. The reason may be attributed to the flexibility KGAGA has in making the dataset to comply with its objective through the formulation of fitness function and a pragmatic constraint of variable coefficients.

A comparative analysis can be conducted with Table 4.6 and Table 4.7, in which D-W test was conducted to check if there is any serious problem of serial correlation arising in the error terms of model fitting. To be more considerate, an ACF test was also conducted to provide more in-depth analysis regarding many inconclusive conclusions in DW test.

Roughly speaking, there is some change concerning the DW test result of model fitting with KGAGA based on different size of samples. The number of cases regarding null hypothesis of no autocorrelation of residuals being rejected in D-W test is 12 in sample 1, while in sample 2, the number reduced to 7. Obviously, the condition of residual autocorrelation improves as the sample size increases for KGAGA. For OLS, the number of cases regarding the same thing is 6 and 7 respectively. The change of sample size doesn't create too much change of DW test results for OLS.

As for ACF test, there are 5 cases failed to pass the test in sample 1, while in sample 2 the number increases to 7 in OLS estimation. There are 10 cases didn't pass ACF test of KGAGA estimation in sample 1, while in sample 2, the number of cases remained the same. It seems that, by and large, the change of sample size doesn't have significant impact in the test results of ACF in model fitting.

Table 4.6. Autocorrelation Test of residuals derived from KGAGA and OLS based on sample 1

item	p num	OLS			KGAGA		
		d value	DW test	ACF	d value	DW test	ACF
1	6	1.503	inconclusive	O	1.223	inconclusive	O
2	3	1.726	O	O	1.534	inconclusive	O
3	4	1.518	inconclusive	O	1.317	inconclusive	O
4	4	1.953	O	O	1.662	inconclusive	O
5	4	2.369	O	O	1.81	O	O
6	5	1.845	O	O	1.813	O	O
7	4	1.737	O	O	0.664	X	X
8	5	1.182	inconclusive	X	0.968	X	inconclusive
9	5	1.322	inconclusive	O	0.538	X	X
10	6	1.333	inconclusive	O	1.245	inconclusive	O
11	5	1.015	inconclusive	O	1.036	inconclusive	inconclusive
12	5	1.743	inconclusive	O	1.659	inconclusive	O
13	4	1.006	X	X	0.9	X	X
14	5	0.99	X	O	0.883	X	X
15	4	0.903	X	inconclusive	0.762	X	X
16	5	1.21	inconclusive	O	1.13	inconclusive	O
17	5	1.788	O	O	1.575	inconclusive	X
18	4	1.358	inconclusive	O	1.309	inconclusive	O
19	5	1.723	O	O	0.685	X	X
20	5	1.338	inconclusive	X	1.455	inconclusive	O
21	5	1.026	X	O	0.639	inconclusive	X
22	5	2.097	O	O	2.052	O	O
23	5	0.993	X	O	0.756	X	X
24	5	1.528	inconclusive	O	1.23	inconclusive	O
25	5	0.744	X	X	0.701	X	X
26	4	1.202	inconclusive	O	1.079	X	O
27	4	1.324	inconclusive	O	1.241	inconclusive	O
28	6	1.574	inconclusive	X	0.93	X	inconclusive
29	6	1.888	O	O	1.899	O	O
30	5	1.625	inconclusive	O	1.549	inconclusive	O

Table 4.7. Autocorrelation Test of residuals derived from KGAGA and OLS based on sample 2

item	p num	OLS			KGAGA		
		d value	DW test	ACF	d value	DW test	ACF
1	6	1.513	inconclusive	O	1.488	inconclusive	O
2	4	1.693	O	O	1.131	O	O
3	4	1.518	inconclusive	O	1.34	inconclusive	O
4	4	1.882	O	O	1.623	O	O
5	4	2.113	O	O	1.669	O	O
6	5	1.77	O	O	1.498	O	O
7	4	1.58	inconclusive	O	1.437	inconclusive	X
8	5	0.698	X	X	0.602	X	X
9	5	1.261	inconclusive	O	0.863	inconclusive	X
10	5	1.219	inconclusive	O	1.111	inconclusive	O
11	6	1.532	inconclusive	O	1.009	inconclusive	X
12	6	1.493	inconclusive	O	1.439	inconclusive	O
13	5	1.148	X	X	0.988	X	X
14	5	1.04	X	O	0.952	X	X
15	4	1.48	X	O	0.89	X	O
16	4	1.138	inconclusive	O	1.136	inconclusive	O
17	5	1.48	inconclusive	X	1.24	inconclusive	X
18	4	1.739	O	O	0.484	O	X
19	5	0.697	X	X	0.589	X	X
20	5	1.367	inconclusive	O	1.386	inconclusive	O
21	5	0.992	X	X	0.885	X	O
22	5	2.084	O	O	2.06	O	O
23	5	1.101	inconclusive	O	0.983	inconclusive	O
24	5	0.961	inconclusive	O	0.931	inconclusive	X
25	4	0.852	X	X	0.85	X	O
26	4	1.058	inconclusive	O	0.96	inconclusive	O
27	4	1.261	inconclusive	O	1.215	inconclusive	O
28	6	1.43	inconclusive	X	1.416	inconclusive	O
29	6	1.619	O	O	1.498	O	O
30	6	1.356	inconclusive	O	1.182	inconclusive	O

Unlike multi-collinearity, residual auto-correlation doesn't affect the unbiased estimation of OLS. Check Table 4.6 - 4.7 with Table 4.3- 4.4, it's hard to find any kind of relationship between residual auto-correlation (failure of DW test or ACF test) and abnormal parameters derived.

4.3.3 Re-estimation of parameters of items of multi-collinearity with stepwise OLS

In order to have a deeper understanding of the influence of multi-collinearity upon model fitting, we select items with max VIF more than 5 and conduct parameter estimation again, but with OLS in stepwise mode which, unlike enter mode which doesn’t delete insignificant variables, removes all insignificant variables, just to see what happen under the impact of multi-collinearity.

Items listed in Table 4.8 and Table 4.9 are selected according to a cut point of max VIF > 5 and are re-estimated with OLS in stepwise mode, parameters removed are marked with a “X”, in each item at least one parameter is removed, some items even have 4 parameters removed. Compared with Table 4.3 and Table 4.4, MAPEs listed in Table 4.8 and Table 4.9 are almost a bit larger than that of their counterpart in enter mode, this implies that by removing variables causing multi-collinearity, regression also lost some fitness. .

It’s also important to note that because variable coefficients with large VIF are removed through stepwise regression, few parameters left in Table 4.8 and Table 4.9 are abnormal.

Table 4.8. Parameters of selected items estimated with stepwise OLS in sample 1

item$_i$	**MAPE$_i$**	**price elas.θ_i**	**base sales λ_i**	**P-mix $\mu_{1i(t=1:T)}$**	**P-mix $\mu_{2i(t=1:T)}$**	**P-mix $\mu_{3i(t=1:T)}$**	**P-mix $\mu_{4i(t=1:T)}$**	**P-mix $\mu_{5i(t=1:T)}$**	**P-mix $\mu_{6i(t=1:T)}$**	**decay rate γ_i**
1	0.0531	-3.673	59.740	X	X	--	1.474	--	X	X
10	0.0217	-1.393	397.82	--	--	X	1.788	X	X	0.839
12	0.0304	X	137.55	X	1.413	2.079	--	--	1.354	0.770
13	0.0267	-1.103	294.12	X	1.898	1.674	--	--	X	X
17	0.0403	-4.321	89.300	X	X	X	--	1.872	--	1.000
18	0.0246	-1.812	141.88	1.852	--	--	X	1.725	--	1.000
21	0.0144	X	240.33	**0.708**	X	1.579	--	--	1.170	1.000
24	0.0395	-1.329	88.854	X	X	2.061	--	--	X	X

Table 4.9. Parameters of selected items estimated with stepwise OLS in sample 2

item$_i$	MAPE$_i$	price elas.θ_i	base sales λ_i	P-mix $\mu_{1i(t=1:T)}$	P-mix $\mu_{2i(t=1:T)}$	P-mix $\mu_{3i(t=1:T)}$	P-mix $\mu_{4i(t=1:T)}$	P-mix $\mu_{5i(t=1:T)}$	P-mix $\mu_{6i(t=1:T)}$	decay rate γ_i
1	0.0569	-3.672	60.160	X	X	--	1.490	--	X	X
7	0.0377	X	83.680	X	--	2.951	--	1.539	--	X
12	0.0304	X	110.39	X	1.413	2.323	--	--	1.354	0.770
13	0.0290	-1.225	290.91	X	1.917	1.846	--	--	X	X
18	0.0304	-4.568	147.23	1.784	--	--	**0.431**	X	--	1
21	0.0242	-1.159	211.24	X	X	1.587	--	--	X	1
24	0.0427	-1.358	88.234	X	X	1.640	--	--	X	1
25	0.0303	-0.939	151.56	--	X	1.696	--	--	X	X
28	0.0263	-1.654	95.300	1.507	X	1.274	--	1.342	X	0.787
29	0.0261	-2.297	118.04	1.344	1.302	1.251	--	X	X	0.786

4.3.4 Comparing accuracy of forecasting based on parameters derived from two different estimators

To investigate forecasting accuracy of our regression model, considering contextual information like price promotion, non-price promotions and decay effect of price discount, estimated by KGAGA and by OLS in enter mode with SPSS, as well as three other forecasting methods capable of multi-step prediction like ARIMA, Exponential Smoothing and Naïve considering cycle length, all of these forecasts are made without any adjustment. With sample 1 and sample 2, the forecasting performance of weekly sales, in terms of MAPE, from the regression model, estimated with two different kinds of methods can be shown in Table 4.10 and Table 4.11, respectively.

The superiority of the forecasting performance of model estimated with KGAGA over that of its counterpart estimated with OGA and OLS is quite obvious. Only 4 cases out of 30 forecasted by re-composition of parameters estimated by KGAGA have higher MAPE than those of its counterpart in forecasts made with parameters estimated in sample 1, check Table 4.10. In sample 2, check Table 4.11, only 10 cases out of 30 did forecasts made by model estimated by KGAGA perform worse than those made by model estimated with OLS in enter mode of SPSS.

Besides, we conducted a paired-samples T test between MAPE of forecasts based on parameters derived from KGAGA and OLS and KGAGA and other forecasting alternatives like ARIMA, ES, and Naïve on sample 1 and sample 2, check Table 4.12, t value is -1.915, -1.167, -1.403, and -2.743 respectively for sample 1 and -1.502, -0.609, 1.605, -2.679 respectively for sample 2, under 0.10 significance level (one-tail) the critical point is 1.311, rejects null hypothesis that, in average, the MAPE of OLS, ES, and Naive is less than or equal to equal to that of KGAGA, and supports alternative hypothesis that, in average, the MAPE of OLS, ES, and Naive is larger than that of KGAGA for forecasting based on sample 1. However, compare the forecasting performance of KGAGA and ARIMA, the test result doesn't support there is any significantly better performance for KGAGA than ARIMA, even though the average MAPE of KGAGA looks better.

For forecasting based on sample 2, in average, the test results support the alternative hypothesis that the MAPE of OLS and Naïve is also larger than that of KGAGA. Supplemented with additional data like average MAPE, it's obvious to see that the forecasting accuracy of model estimated with KGAGA is better than that of model estimated by OLS in sample 2.

Table 4.10. Comparison of the accuracy of multi-step forecasting with parameters derived from KGAGA, OLS and forecasting from other methods based on sample 1

Forecasting accuracy measured by MAPE						
item	**KGAGA**	**OGA**	**OLS**	**ARIMA**	**ES**	**NAÏVE**
1	0.243	0.556	0.502	0.531	0.506	0.411
2	0.349	0.349	0.321	0.301	0.312	0.321
3	0.078	0.078	0.078	0.361	0.286	0.079
4	0.336	0.336	0.336	0.269	0.253	0.336
5	0.073	0.073	0.073	0.270	0.294	0.073
6	0.209	0.210	0.214	0.175	0.140	0.214
7	0.254	0.393	0.233	0.176	0.211	0.393
8	0.520	0.602	0.608	0.360	0.288	0.572
9	0.134	0.318	0.338	0.160	0.171	0.349
10	0.213	0.213	0.213	0.381	0.414	0.213
11	0.221	0.221	0.221	0.193	0.135	0.221
12	0.240	0.240	0.240	0.484	0.472	0.240
13	0.247	0.283	0.353	0.393	0.380	0.304
14	0.209	0.209	0.209	0.385	0.399	0.209
15	0.197	0.204	0.194	0.285	0.208	0.194
16	0.186	0.186	0.186	0.309	0.463	0.186
17	0.335	0.609	0.489	0.267	0.200	0.602
18	0.119	0.118	0.200	0.212	0.535	0.225
19	0.302	0.337	0.543	0.233	0.252	0.544
20	0.230	0.230	0.230	0.136	0.241	0.230
21	0.228	0.218	0.228	0.270	0.212	0.224
22	0.305	0.305	0.305	0.809	1.353	0.305
23	0.352	0.352	0.352	0.279	0.234	0.352
24	0.595	0.595	0.595	0.284	0.299	0.595
25	0.106	0.118	0.123	0.199	0.170	0.123
26	0.519	0.527	0.380	0.279	0.393	0.505
27	0.299	0.399	0.214	0.263	0.333	0.311
28	0.257	0.337	0.662	0.213	0.212	0.223
29	0.106	0.237	0.085	0.204	0.128	0.081
30	0.249	0.249	0.213	0.198	0.192	0.249
mean	**0.257**	0.303	0.302	0.296	0.323	0.296
std	0.123	0.148	0.157	0.132	0.221	0.146

Table 4.11. Comparison of the accuracy of multi-step forecasting with parameters derived from KGAGA and OLS on sample 2

Forecasting accuracy measured by MAPE						
item	**KGAGA**	**OGA**	**OLS**	**ARIMA**	**ES**	**NAÏVE**
1	0.321	0.368	0.274	0.464	0.356	0.254
2	0.163	0.254	0.256	0.574	0.744	0.222
3	0.189	0.182	0.180	0.735	0.246	0.173
4	0.188	0.192	0.192	0.319	0.101	0.131
5	0.247	0.247	0.247	0.120	0.426	0.247
6	0.122	0.141	0.125	0.169	0.222	0.171
7	0.227	0.227	0.227	0.583	0.355	0.227
8	0.133	0.137	0.176	0.241	0.302	0.376
9	**0.680**	0.653	**0.842**	0.324	0.280	0.832
10	0.299	0.301	0.304	0.268	0.128	0.294
11	0.916	1.274	0.959	1.340	0.917	2.238
12	0.102	0.241	0.229	0.071	0.080	0.555
13	0.517	0.664	0.429	0.158	0.233	0.415
14	0.175	0.123	0.376	0.081	0.082	0.200
15	**0.846**	0.849	**0.578**	0.167	0.121	0.819
16	**0.437**	1.046	**0.958**	0.146	0.148	0.687
17	0.231	0.248	0.252	0.457	0.188	0.363
18	0.159	0.847	0.210	0.388	0.295	0.521
19	**0.395**	1.182	**0.540**	0.392	0.392	0.648
20	0.181	0.269	0.170	0.370	0.329	0.148
21	**0.240**	0.204	**0.370**	0.477	0.329	0.438
22	0.338	0.189	0.305	0.226	0.285	0.177
23	0.268	0.208	0.238	0.306	0.271	0.221
24	0.196	0.115	0.102	0.299	0.240	0.275
25	**0.286**	0.322	**0.381**	0.155	0.141	0.319
26	0.334	0.294	0.351	0.318	0.366	0.652
27	0.357	0.411	0.357	0.307	0.297	0.658
28	0.284	0.291	0.289	0.240	0.252	0.249
29	0.178	0.137	0.166	0.213	0.209	0.325
30	0.241	0.287	0.218	0.258	0.257	0.383
mean	**0.308**	0.397	0.343	0.339	**0.286**	0.441
std	0.200	0.327	0.227	0.245	0.176	0.396

Table 4.12. T test results for forecasting accuracy comparison between KGAGA and other alternatives

comparison type	Paired sample T test(one-tail), α=0.10			
	Sample 1		Sample 2	
	t value	result	t value	result
KGAGA vs OGA	-3.021	X	-2.156	X
KGAGA vs OLS	-1.915	X**	-1.502	X
KGAGA vs ARIMA	-1.167	O	-0.659	O
KGAGA vs ES	-1.403	X	1.605	X*
KGAGA vs Naive	-2.743	X	-2.679	X

*null hypothesis is $MAPE_{KGAGA} \geq MAPE_{ES.}$

** X denotes rejection of null hypothesis.

4.3.5 Implications for forecasters

Multi-collinearity, a very common issue in linear regression, has a deep and wide negative effect upon the model building process in which variable selection, and model fitting, even model specification are all affected by OLS regressor to such an extent that sporadic variables become insignificant and are deleted, parameters estimated are distorted and become unrealistic and incorrect eventually, check Table 4.1 and Table 4.2 as well as Table 4.3 and Table 4.4. As a result, statistical inference made from the model seems meaningless, out-of-sample forecasts made by the model become out of track, check Table 4.10-11, all these phenomenon can be attributed to the lack of enough variations in the data pertained to certain variables.

On the other hand, if KGAGA proposed in this study is used instead, the issue of insignificant sporadic variable will not be present, in particular, estimation of model parameters will be more realistic and correct, check Table 4.1 and Table 4.2, and these results will be of great help in improving the performance of out-of-sample forecasting eventually, check Table 4.10 and Table 4.11. The method of re-composition of effect parameters estimated via KGAGA according to anticipated variations in the context like promotions and decay effect of price promotions, which can be easily checked in the promotion proposals and the historical data, is a practical method of both efficiency and effectiveness, even though it may be not as handy and efficient as ARIMA or ES at present.

5. The Contextual Forecasting Model

The formulation of contextual forecasting model in this thesis is based on the following assumptions:

Assumptions

1. Cross-brand effects of competitors' competitive actions are roughly constant in the same period across years, and will remain so in the immediate future.
2. The reference price and price discounts of the same item are the same for all the stores under one retailer in the same period.
3. The effects of non-price promotion mixes held in the stores under one retailer are homogenous.
4. The weekend just cover Saturday and Sunday, Friday's evening is not included.
5. The weekend effect is roughly fixed across products and across weeks.
6. For the three groups of indicator parameters shown below, at most one condition arises in each group at a time.
7. Normally, a trading year is composed of *52* weeks.
8. The decay rate of similar price discount effect in promotion sessions remains constant for the same item and will remain so in the forecasting horizon.

5.1 Formulation of a Regression Model of Promotion Effects and Holiday Effects

The first objective in our forecasting model involves decomposing the promotional sales of products of a company into simple components easy to handle. Eq. (5.1) of our regression model is motivated by Dick R. Wittink et al.'s analytical models in (Foekens, Leeflang, and Wittink, 1999; Heerde, Leeflang, and Wittink, 2002a; 2002b). The model can be formulated as

$$S_{it} = \lambda_i \left(P_{it} / \widehat{P}_i\right)^{\theta_i} \prod_{l=1}^{n} \mu_{l_i}{}^{D_{lit}} \beta_i{}^{E_{it}} \prod_{r=1}^{o} w_{ri}{}^{H_{rit}} \ell^{\varepsilon_{it}} \quad , \ \forall \ t \in Q \qquad (5.1)$$

Where, most of the variables and parameters in EQ. (5.1) are already explained in section 3, whereas new ones in Eq. (5.1) can be defined as follows:

H denotes an indicator parameter of holiday.

H_{rit} is the r-th component of a vector of o indicator parameters of holiday (H_{1it}, H_{2it} , ... , H_{oit}) to indicate whether there is any holiday(s) in a certain period t or not. $H_{rit} = 1$ denotes a holiday of type r arises, the default value of $H_{rit} = 0$.

w_{ri} denotes holiday effect parameter of holiday type r in period t of item i.

Besides, additional notations listed below may be helpful in the following sections.

R denotes the set of mixed periods in which there are more than one promotion held in the same period.

φ denotes the weekend effect, which is derived via KGAGA based on data in mixed periods.

$d(t_1)$ denotes the length of sub-period t_1 of t, $t \in Z$. $0 \leq d(t_1) \leq 7$.

$d(t_2)$ denotes the length of sub-period t_2 of t, $t \in Z$. $0 \leq d(t_2) \leq 7$. $d(t_1)+d(t_2)= d(t)$, because in a week there are at most two different kinds of promotion mixes held.

δ denotes the duration of the weekend.

δ_{t1} denotes the duration of weekend covered by sub-period t_1.

Take natural logarithm in both sides of Eq. (5.1), we get the following:

$$\ln S_{it} = \ln \lambda_i + \theta_i \ln\left(P_{it} / \widehat{P}_i\right) + \sum_{l=1}^{n} D_{lit} \ln \mu_{li} + E_{it} \ln \beta_i + \sum_{r=1}^{o} H_{rit} \ln \omega_{ri} + \varepsilon_{it}, \quad \forall t \in Q \tag{5.2}$$

Thus, a nonlinear model like Eq. (5.1) is transformed to a linear regression model, see Carroll and Ruppert (pp.115-160,1988); and Franses and McAleer (1998), which is the underlying model to conduct model fitting and model checking in this thesis.

5.2 Parameter Estimation with KGAGA

To take into account of all the influential and sporadic cue factors in various sub-periods of the training period, the number of variables may amount to such a quantity that conventional parameters estimation method like ordinary least square, maximum likelihood method, and so on may become incompetent, due to the issue of collinearity (Belsley, 1982; Belsley, Kuh, and Welsch, 1980; Hendry, 2000), insignificant parameters (Bunn and Salo, 1996; Hendry, 2000), or small sample size. Therefore, in this thesis we use KGAGA which could estimate parameters effectively and efficiently as mentioned in section 3.

5.3 The Underlying Nearest Neighbor Method

Usually in the same period of time in lunar calendar between different years, the feature of promotion held is very similar to each other, this makes the nearest neighbor method an ideal underlying principle of initial forecasting and subsequent mechanical adjustment in our thesis. Generally speaking, the effect of each promotion mix in promotion sessions of last year is estimated with KGAGA, the same or the most similar estimated effect of promotion mix is used to forecast next year's sales in promotion session within forecasting horizon.

The nearest neighbor model for prediction of avalanches is first proposed by Buser (1983), which selects ten days most similar to the given situation from 20-years data. It has the advantage of simple computational approach. The model represents only the most similar situations corresponding to the current situation and the decision making itself depends on the forecaster.

Felber and Bartelt (2003) argue that nearest neighbor models attempt to compare similar situations in the past with current data and assume that similar events are likely to occur in similar conditions.

The nearest neighbor method works by selecting geometric segments in the past of the time series similar to the last segment available before the observation we want to forecast (see, e.g., Stone, 1977; Cleveland, 1979 and HaÈrdle and Linton, 1994).

Meade (2002) notes that the intuition underlying nearest neighbor method is that if a previous pattern can be identified as similar to the current behavior of the time series, then the previous subsequent behavior of the series can be used to predict behavior in the immediate future.

5.4 The Mechanical Adjusting Mechanism

The mechanical mechanism of this thesis stresses that adjustments of forecasting are based on the anticipated changes of the context of promotions and moving holidays in the forecasting periods, which can not be handled by the regression model alone, in this thesis a set of equations are formulated to do this job, they are natural extension of the model. The objective resides in improving the performance of the model, making our final forecasts better reflect these changes in prospect.

The weekend effect is a critical consideration in our mechanical adjusting mechanism, because in weekly sales of CPG products, the sales volume in the weekend usually is about the equal of the combined sales volumes of the five weekdays or even more than the latter, see Singh and Takhtani (2005). Martinez-Ruiz et al. (2006) also investigated promotional effects during the weekend and concluded that for all 6 brands, the weekend in promotional periods has a positive and differential impact over sales of coffee.

To adequately assess the combined effect of promotion in a specific week of the forecasting horizon and thus make adjustment accordingly, in a week because some promotion sessions cover only one day of weekend, others cover two days of weekend, the impact of this difference is not trivial because of combined multiplicative effects of promotion and the weekend, it makes sense to take the weekend effect into account.

In this thesis, we use KGAGA, check Figure 5.1, to estimate the weekend effect of CPG products of company A and get an average effect of 2.873 for per day of weekend.

5.4.1 Seasonal index Realignment (SIR)

Based on domain knowledge, sales volume of the last week or average sales volume of the last few weeks (adjusted with calendar effect) in the reference periods is a better predictor to sales of the first few weeks in the forecasting periods next year than sales of the same weeks in the referenced periods in Taiwan, check Figures. 5.1-5.2. Thus, the corresponding formula can be formulated as

$$\hat{S}_{it'} = \eta_i \pi_{it(t=\Omega)} e_{1it'} e_{2it'} e_{3it'} e_{4it'} e_{5it'(t'=t+52-ws)} \qquad ,\forall t' \in Z \tag{5.3}$$

Eq. (5.3) is an example of modified Naïve method used for multiple-step out-of-sample forecasting considering cycle length which is a year. In which, $\eta_i \pi_{it(t=\Omega)}$ stands for normal (baseline) sales of the last week(s) in the training periods and is the most recent related data available. While $e_{5it'}$ stands for pre LNY (Lunar New Year) effect arises annually in a period of around 4 weeks right before LNY. In this period, sales volumes are usually much higher than usual even without any promotion.

Since pre LNY effect has not been incorporated as a variable into our regression model for parsimonious purpose, nevertheless, it will dominate the seasonal indices in corresponding periods, hence, it's quite intuitive to use these indices as proxy variable of pre LNY effect, denoted as $e_{5it'}$. Because there is usually a time shift of the timing of LNY from year to year, to forecast unit sales of weeks after LNY in next year, the week number referenced corresponding to the week in the forecasting horizon has to be adjusted.

Let LNY(*t'*) denote the week number of LNY in the year to be forecasted, LNY(*t*) denote the week number of LNY in the referenced year, then, let *ws* = LNY(*t'*) – LNY(*t*) represents the number of weeks shifted between two different years as the week of LNY is concerned. If *ws* > 0, it means the sequence of week of LNY(*t'*) in the forecasting year will be *ws* weeks later than that of LNY(*t*) in the referenced year; on the other hand, if *ws* < 0, it means the sequence of week of LNY(*t'*) in the forecasting year will be *ws* weeks earlier than that of LNY(*t*) in the referenced year. Therefore, the most right term in Eq. (5.3) $e_{5it'}$ will be replaced by $\pi_{i(t-ws)}$. Thus Eq. (5.4) will become

$$\hat{S}_{it'} = \eta_i \pi_{it(t=\Omega)} e_{1it'} e_{2it'} e_{3it'} e_{4it'} \pi_{i(t-ws)} \qquad , \forall t' \in Z \tag{5.4}$$

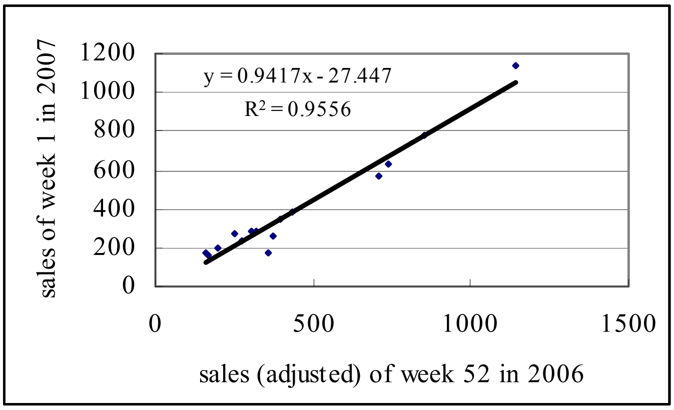

Figure 5.1. Regression of week 52's sales (2006) to sales of week 1 in 2007.

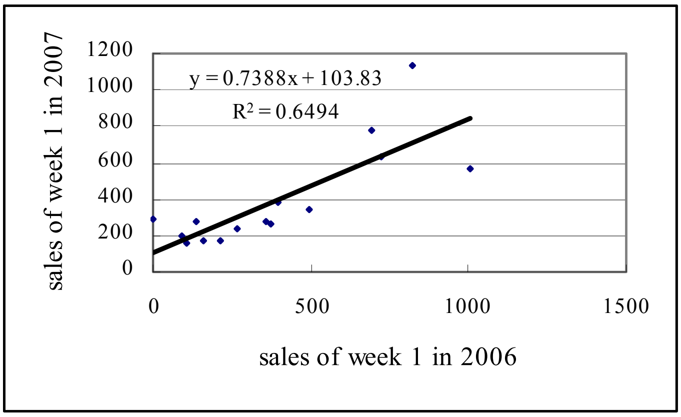

Figure 5.2. Regression of week 1's sales (2006) to sales of week 1 in 2007.

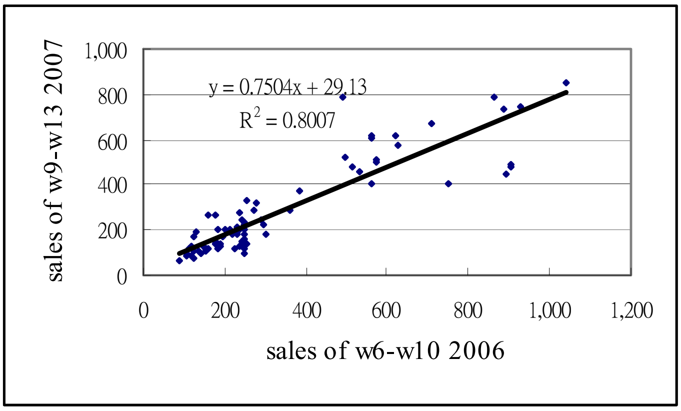

Figure 5.3. Regression of sales of w6-w10 2006 to that of w9-w13 in 2007.

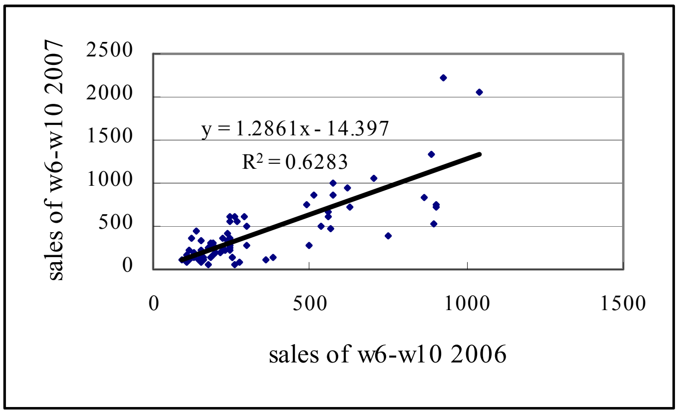

Figure 5.4. Regression of sales of w6-w10 2006 to that of w6-w10 in 2007.

We also find that sales of the same weekly order as that in a different year after LNY are more aligned than weekly sales of the same ordinary sequential order between different calendar years, check Figures 5.3-5.4, in which R^2 from sales of weeks after LNY (in week 8) in 2007 regressed against those of weeks after LNY (in week 5) in week of 2006 is 0.8, much better than the ordinary week n corresponding to the same week n, n = 1,2,..,5, regression between different years, which only has a R^2 of 0.628, please check Figure 5.4.

Based on the finding mentioned above, to forecast the sales of weeks after the week of LNY in a new year, denoted as $\hat{S}_{it'}, t' > LNY(t')$, Eq. (5.5) can be of use, which is modified from Eq. (5.4) :

$$\hat{S}_{it'} = \eta_i \pi_{i(t-ws)} e_{1it'} e_{2it'} e_{3it'} e_{4it'}, \qquad t' > LNY(t') \qquad , \forall t' \in Z \tag{5.5}$$

5.4.2 Proportional adjustment (PA)

Quite often, in the week of LNY, there is a small part of pre LNY present prior to the eve of LNY, or the last week of pre LNY is mixed with a small part of LNY in the referenced period, but the condition of the corresponding week in forecasting horizon is different, in these cases, to get a proper estimation of these effect multipliers of calendar effect in the forecasting periods, we must get them restored to regular ones (a whole week only covered by purely one kind of holiday related effect like pre LNY effect or holiday effect of LNY) first, and then proceed to calculate the changed mixed effect in the forecasting period.

The adjusting equations are used to calculate the mixed effect of pre LNY and LNY present in the same week:

$$e_{4it} = \frac{(d(t_1) - d(\delta_{t_1}) + d(\delta_{t_1})\varphi) e_{5it_1} + (d(t_2) - d(\delta_t) + d(\delta_{t_1}) + (d(\delta_t) - d(\delta_{t_1}))\varphi) e^*{}_{4it_2}}{d(t) + d(\delta_t)(\varphi - 1)},$$

$$\forall t \in (Q \cup R) \tag{5.6}$$

In Eq. (5.6), the part $(d(t_1) - d(\delta_{t_1}) + d(\delta_{t_1})\varphi) e_{5it_1}$ represents the sum of the effect of the last week in pre LNY in the duration of weekdays covered by sub-period t_1 and the effect of the last week in pre LNY in the duration of weekend covered by sub-period t1 times the weekend effect. While the term $(d(t_2) - d(\delta_t) + d(\delta_{t_1}) + (d(\delta_t) - d(\delta_{t_1}))\varphi) e^*{}_{4it_2}$ stands for the sum of the effect of regular LNY in the duration of weekdays covered with sub-period t_2 in the referenced period and the effect of regular LNY in the duration of weekend covered by sub-period t_2 in the referenced period times the weekend effect. Every parameter in Eq. (5.6) except e^*_{4it} is known, so e^*_{4it} can be obtained.

Note that the daily effect of each normal weekday in a week is assumed to be 1. Eq. (5.6) actually is a daily effect weighted average formula of mixed weekly effect of pre LNY and LNY present in the same week.

It follows that the mixed effect in the forecasting period ($e_{4it_1'}$) could be computed through the following formula:

$$e_{4it'} = \frac{(d(t_1') - d(\delta_{t_1'}) + d(\delta_{t_1'})\varphi)e^*_{4it_1'} + (d(t_2') - d(\delta_t) + d(\delta_{t_1'}) + (d(\delta_t) - d(\delta_{t_1'}))\varphi)e_{rit_2'}}{d(t) + d(\delta_t)(\varphi - 1)} \ ,$$

$$\forall t' \in Z \qquad (5.7)$$

Where, erit2’ may be effect of the last week in pre LNY or just base effect equal to 1. In Eq. (5.7), the part $(d(t_1') - d(\delta_{t_1'}) + d(\delta_{t_1'})\varphi)e^*_{4it_1'}$ stands for the sum of the effect of the regular LNY in the duration of sub-period $t_1{'}$ in the weekdays in the forecasting period and the effect of the regular LNY in the duration of sub-period $t_1{'}$ in the weekend times the weekend effect. While the part of $(d(t_2') - d(\delta_t) + d(\delta_{t_1'}) + (d(\delta_t) - d(\delta_{t_1'}))\varphi)e_{rit_2'}$ represents the sum of the effect of the last week in pre LNY in the duration of sub-period t_2’ in the weekdays in the forecasting period and the effect of the last week in pre LNY in the duration of sub-period $t_2{'}$ in the weekend times the weekend effect.

In the same token, regular effect of the last week in pre LNY ($e^*_{5it_1'}$) in the

$$e_{5it} = \frac{(d(t_1) - d(\delta_{t_1}) + d(\delta_{t_1})\varphi)e^*_{5it_1} + (d(t_2) - d(\delta_t) + d(\delta_{t_1}) + (d(\delta_t) - d(\delta_{t_1}))\varphi)e_{4it_2}}{d(t) + d(\delta_t)(\varphi - 1)},$$

1’

$$\forall t \in (Q \cup R) \qquad (5.8)$$

$$e_{5it'} = \frac{(d(t_1') - d(\delta_{t_1'}) + d(\delta_{t_1'})\varphi)e^*_{5it_1'} + (d(t_2') - d(\delta_{t'}) + d(\delta_{t_1'}) + (d(\delta_{t'}) - d(\delta_{t_1'}))\varphi)e_{4it_2'}}{d(t) + d(\delta_{t'})(\varphi - 1)} \ ,$$

$$\forall t' \in Z \qquad (5.9)$$

Quite often, over time the level of baseline sales changes owing to fundamental variation of economic growth, new competitor, competitor’s new product, competitor’s new promotion, and so on, in such case, average normal sales (η_i) in Eq. (5.3) will not be adequate anymore to be used for forecasting purpose, instead, an average of normal sales in most recent weeks of the training period can be a better estimate of average normal sales, therefore, Eq. (5.3) can be modified to the following:

$$\hat{S}_{it'} = \eta_{i*}\pi_{i(t-ws)}e_{1it'}e_{2it'}e_{3it'}e_{4it'}, \quad t' > LNY(t') \quad ,\forall t' \in Z \qquad (5.10)$$

Where, η_{i*} denotes an average of most recent 3 weeks' sales of item i in the training period. As a general rule in this thesis, if η_{i*} is more than 5% in deviation from η_i in Eq. (5.3), Eq. (5.10) will be used to recombine parameters for forecasting purpose.

5.4.3 Total adjustment (TA)

As the combination of both adjustments of SIR and PA, TA offers the most comprehensive adjustment, including seasonal index realignment after LNY and proportional adjustment of promotional effect and holiday effect as well as weekend effect and a level adjustment mentioned above, in this thesis.

6. Empirical Study—part 2

6.1 The Background of Our Empirical Study

This study has a focus on the adjustment of model-based forecast of weekly unit sales of several series of CPG products, manufactured by Company A, under retailer B. Company A is a leading manufacturer specialized in dehumidifier and deodorizer products in Taiwan. While retailer B is an international outlet of DIY products.

A sales data set of 30 items in 2007, aggregated from retailer B's 26 outlets, coupled with price promotion, non-price promotion data, as well as promotion proposals, which were set up in 2007, of the first 4 months in 2008, are used to conduct our empirical study. The training dataset covers two periods, the first period covers the whole year of 2007, the dataset of this period can be denoted as sample 1, forecasting horizon is the first 6 weeks of 2008. The second period ranges from the beginning of 2007 to the 10th week of 2008, the training data of this period can be denoted as sample 2, forecast horizon ranges from 11th week to 16th week of 2008. To show the volatility of sales across weeks, related statistical characteristics of sales data in sample 1 and 2 are displayed in Table 6.2 and Table 6.3 respectively. Note that if standard deviation of sales is more than half of the average sales of any item in both tables, it was marked with bold surface.

The underlying equation used in model fitting was Eq. (6.2), all the effect parameters in Eq. (6.2) were estimated through KGAGA with objective function set as Eq. (3.4) and constraints set realistically from contextual knowledge, such as price elasticity parameter to be in the range of [-9, 1.0], non-price promotion effect multiplier to be in the range of [1, 4], holiday effect multiplier to be in the range of [0.5, 3]. Besides, the number of types of non-price promotion mixes n in Eq. (6.2) was set to be 6, therefore, there are about 6+1 = 7 types of different combination of promotion activities across seasons. The decay rate of price promotions is set in the range of [0.3, 1]. In addition, the number of moving (lunar) holiday type o was 4, that means there are 4 main different types of moving holidays each year, to reflect the actual business settings. KGAGA programs were run with Matlab 7.1 with population size set at 1000, the number of generations to terminate the program is set at 200, total time of a run by a desktop pc takes about 200 seconds in average.

Effect parameters estimated with KGAGA and mixed effect parameters reassessed in the mixed periods on both sample 1 and sample 2 were recomposed according to the expected variations of promotions in the promotion proposal as initial forecasts without any adjustment. The multiple-step out-of-sample forecasts with ARIMA and ES were run with SPSS 13. In busy season, because of intensive promotion campaigns, ARIMA and ES tends to underestimate weekly unit sales, the rule of adjustment for forecasts of ARIMA can be formulated as

$$\text{ARIMA ad} = \text{forecast of ARIMA} * 1.2 \tag{6.1}$$

$$\text{ES ad} = \text{forecast of ES} *1.2 \tag{6.2}$$

However, in time of off season, ARIMA and ES tend to overestimate unit sales, the rule of adjustment can be formulated as

$$\text{ARIMA ad} = \text{forecast of ARIMA} * 0.8 \tag{6.3}$$

$$\text{ES ad} = \text{forecast of ES} * 0.8 \tag{6.4}$$

Thus, the performance of various adjustment methods proposed in this study can be compared with their counterpart of ARIMA and ES.

6.2 The Design of Our Empirical Study

In order to take both the busy season and off season into account to have a proper assessment of the performance of different adjustment methods, the forecasting horizon is designed to consist of two periods of equal duration, the first period includes the first 6 weeks of 2008 which covers the busiest season, ie, the LNY season in Taiwan, and can be denoted as busy season, while the second one starts from the 11th week and ends at the16th week of 2008, which is one of the off seasons in the same year and can be denoted as off season.

As the forecasting target is concerned, 30 items of products were selected to conduct our empirical research. The relevant prices and promotion activities can be found in promotion proposals which actually are the source of anticipated variations in promotions. Another source of anticipated variations in calendar effects is the calendar.

To properly evaluate the performance of various methods in adjusting original forecasts of the model, which can be denoted as NA, made by regression model estimated with KGAGA, SIR was performed first to adjust NA, followed by PA to adjust the same NA. Then, the combination of SIR and PA were used to adjust NA, denoted as TA. The busy season was the first forecasting horizon, and off season was the second forecasting horizon. In addition, for the purpose of comparative reference, forecasts with Box and Jenkins ARIMA and ES were derived, adjustments of forecasts from ARIMA and ES were derived with Eq. (6.1) and Eq. (6.3) in busy season, and Eq. (6.2) and Eq. (6.4) in off season, respectively.

6.3. Empirical Results

6.3.1 The results of model fitting

The average error of model fitting in terms of MAPE in general for KGAGA is below 5%. Most parameters derived from KGAGA are consistent with our expectations. In contrast, average MAPE for OLS is a bit larger than that of KGAGA in both samples, besides, there are many parameters derived which seems to be unreasonably low or unreasonably high which are marked with boldface in item 1, 8, 9, 14, 16, 17, 21, 23, 24, 26, and item 27 in sample 1; item 8, 9, 14, 17, 21, 23, 25, 26, and item 27 in sample 2, check Table 6.2 and 6.3. These abnormalities deserve more in-depth investigation.

6.3.2 The results of model checking

The normality test, consisting of one-sample Kolmogorov-Smirnov test and Q-Q plot, in which, this model passed the test with data from sample 1 or sample 2 without any problem based on standardized error term ε_{it} in Eq. (6.2) and natural logarithm of predicted weekly unit sales denoted as $\ln \widetilde{S}_{it}$ in Eq. (6.3).

As collinearity is concerned, with one more category of indicator variables considering moving holiday effects in the new model in section 4, max VIF of regression for each item doesn't change too much from that of the original model, except a few items like item 7 and item 15. Detailed results of multi-collinearity analysis in terms of VIF can be shown in Table 6.8, note that those max VIFs more than 5 are marked with boldface.

Table 6.1. Summary of promotion proposals for year 2007 and part of 2008 of Company A's products under retailer B

Promotion sessions with content denoted as ($P_{it} / \widehat{P}_i , D_l$)

item (i)	Product type	2007									2008	
		12/29-2/28	3/29-4/24	4/26-6/12	6/14-8/07	8/09/9/11	9/13-11/13	11/15-12/15	10/01-12/31	12/27-2/12	2/14-4/1	4/3-4/29
1	Insect pest	195/219,D_3	1,D_0	195/219,D_0	1,D_0	189/219,D_0	189/219,D_5	195/219,D_5	189/219,D_6	169/219,D_3	1,D_0	195/219,D_0
2	cleaner	99/109,D_3	99/109,D_0	99/109,D_0	99/109,D_1	99/109,D_1	99/109,D_1	99/109,D_0	99/109,D_5	99/109,D_3	1,D_0	99/109,D_1
3	cleaner	90/109,D_4	90/109,D_0	90/109,D_0	90/109,D_1	89/109,D_1	1,D_5	90/109,D_5	1,D_5	89/109,D_4	1,D_0	90/109,D_0
4	Insect pest	85/89,D_4	1,D_0	85/89,D_0	85/89,D_0	85/89,D_1	85/89,D_5	1,D_0	85/89,D_5	85/89,D_3	1,D_0	85/89,D_0
5	clear dose	90/99,D_3	1,D_0	1,D_0	1,D_0	85/89,D_1	1,D_5	90/99,D_5	1,D_5	90/99,D_3	1,D_0	1,D_0
6	mop	199/229, D_3	1,D_0	199/229, D_0	1,D_2	1,D_1	1,D_2	199/229, D_5	1,D_5	189/229, D_3	1,D_2	199/229, D_0
7	mop	50/60, D_3	1,D_0	1,D_0	1,D_0	1,D_1	1,D_1	50/60, D_5	1,D_5	55/60, D_3	1,D_0	1,D_0
8	dehumidifier	52/59, D_3	1,D_1	52/59, D_1	52/59, D_2	1,D_1	1,D_1	52/59, D_5	1,D_6	49/59, D_3	52/59, D_0	52/59, D_0
9	dehumidifier	75/89,D_4	1,D_0	1,.D_0	75/89,D_0	1,D_1	1,D_1	1,D_5	1,D_6	1,D_3	1,D_0	79/89,D_5
10	cleaner	49/59,D_4	1,D_5	49/59,D_0	49/59,D_2	1,D_0	1,D_5	49/59,D_6	49/59,D_6	49/59,D_3	1,D_0	49/59,D_5
11	aromatizer	329/359,D_3	1,D_1	329/359,D_1	329/359,D_1	329/359,D_2	1,D_1	329/359,D_5	329/359,D_5	329/359,D_3	329/359,D_0	329/359,D_0
12	mop	50/60, D_3	50/60, D_0	50/60, D_0	1,D_1	1,D_1	1,D_2	44.5/60, D_6	44.5/60, D_6	50/60, D_3	1,D_0	1,D_0
13	aromatizer	52/65, D_3	1,D_1	49/65, D_1	52/65, D_1	1,D_0	49/65, D_1	44.5/65, D_6	44.5/65, D_6	49/65, D_3	49/65, D_1	55/65, D_0
14	mop	189/219, D_3	189/219, D_1	189/219, D_1	189/219, D_1	189/219, D_3	189/219, D_3	189/219, D_6	189/219, D_6	189/219, D_3	189/219, D_2	1,D_0
15	cleaner	100/145, D_4	1,D_0	1,D_0	1,D_1	100/145, D_2	1,D_0	1,D_0	1,D_0	100/145, D_4	1,D_0	1,D_0
16	cleaner	49/59,D_4	49/59,D_0	1,D_0	45/59,D_1	1,D_0	1,D_0	44.5/59,D_6	44.5/59,D_6	49/59,D_4	1,D_0	1,D_0
17	deodorizer	89/95,D_3	1,D_1	1,D_0	1,D_1	89/95,D_1	89/95,D_2	89/95,D_5	89/95,D_5	75/95,D_3	89/95,D_1	89/95,D_1
18	cleaner	42/55,D_4	45/55,D_5	1,D_1	1,D_1	1,D_0	1,D_5	1,D_1	1,D_5	42/59,D_4	1,D_0	49/59,D_1
19	Insect pest	65/79,D_3	1,D_1	1,D_0	1,D_1	1,D_0	1,D_1	65/79,D_5	1,D_1	1,D_3	1,D_0	65/79,D_1
20	camphor oil	1,D_3	89/129,D_3	1,D_2	1,D_6	1,D_2	1,D_1	1,D_0	1,D_1	1,D_3	1,D_0	1,D_2
21	deodorizer	59/65,D_3	1,D_0	1,D_0	59/65,D_2	1,D_0	1,D_0	1,D_0	49.5/65,D_6	55/65,D_3	1,D_1	1,D_0
22	deodorizer	59/65,D_3	1,D_0	1,D_0	59/65,D_2	1,D_0	1,D_0	1,D_0	49.5/65,D_6	55/65,D_3	1,D_0	1,D_0

Promotion sessions with content denoted as ($P_{it}/\widehat{P}_i, D_l$)

item	Product type	2007									2008	
(i)		12/29-2/28	3/29-4/24	4/26-6/12	6/14-8/07	8/09/9/11	9/13-11/13	11/15-12/15	10/01-12/31	12/27-2/12	2/14-4/1	4/3-4/29
23	deodorizer	$1,D_3$	$1,D_0$	$1,D_0$	$1,D_2$	$1,D_1$	$1,D_1$	$1,D_5$	$1,D_2$	$1,D_3$	$45/49,D_0$	$45/49,D_0$
24	mop	$50/60, D_3$	$50/60, D_1$	$50/60, D_1$	$1,D_1$	$1,D_1$	$1,D_2$	$44.5/60,D_6$	$44.5/60,D_6$	$50/60, D_3$	$1,D_0$	$1,D_0$
25	deodorizer	$59/65,D_3$	$1,D_0$	$1,D_0$	$59/65,D_2$	$1,D_0$	$1,D_0$	$49.5/65,D_6$	$49.5/65,D_6$	$55/65,D_3$	$1,D_0$	$1,D_0$
26	deodorizer	$85/95,D_3$	$1,D_2$	$1,D_0$	$1,D_2$	$85/95,D_1$	$1,D_0$	$85/95,D_1$	$1,D_1$	$79/95,D_3$	$85/95,D_0$	$1,D_0$
27	deodorizer	$85/95,D_3$	$1,D_2$	$1,D_0$	$1,D_2$	$85/95,D_1$	$1,D_0$	$85/95,D_1$	$1,D_1$	$79/95,D_3$	$85/95,D_0$	$1,D_0$
28	deodorizer	$1,D_3$	$1,D_0$	$1,D_1$	$89/99,D_2$	$1,D_1$	$85/99,D_5$	$79.5/99,D_5$	$79.5/99,D_6$	$85/99,D_3$	$1,D_0$	$89/99.D_6$
29	deodorizer	$1,D_3$	$1,D_0$	$1,D_0$	$89/99D_2$	$1,D_0$	$85/99,D_5$	$79.5/99,D_5$	$79.5/99,D_6$	$85/99,D_3$	$1,D_0$	$89/99,D_2$
30	dehumidifier	$1,D_3$	$1,D_0$	$1,D_0$	$119/138,D_2$	$119/138,D_2$	$1,D_5$	$99.5/138,D_6$	$99.5/138,D_6$	$1,D_4$	$1,D_0$	$1,D_5$

Notes: $P_{it}/\widehat{P}_i$ denotes the discount price rate.
D_l, $l = 0{:}6$, denotes specific type of promotion mix as explained in section 3.1

Table 6.2. Statistical characteristics of sales data in sample 1

item	1	2	3	4	5	6	7	8	9	10
min	37	24	120	124	60	68	45	78	270	239
max	249	262	413	622	245	299	350	298	938	1135
avg	95.87	100.56	212.62	257.48	119.40	135.67	121.94	169.37	547.21	534.15
s.d.	44.66	45.03	57.00	103.21	36.98	40.07	69.36	50.19	155.95	197.82

item	11	12	13	14	15	16	17	18	19	20
min	94	90	209	140	487	84	57	116	103	84
max	299	435	832	745	2227	318	351	462	540	717
avg	178.96	175.37	449.69	315.10	868.08	175.17	132.65	242.94	220.62	265.96
s.d.	40.94	68.86	134.96	114.39	311.54	55.92	61.64	85.60	97.73	142.23

item	21	22	23	24	25	26	27	28	29	30
min	139	68	52	54	93	47	43	71	75	90
max	561	332	293	333	464	414	372	209	279	294
avg	264.71	137.85	125.92	131.38	195.02	120.17	110.73	126.73	161.67	186.85
s.d.	79.65	68.97	60.52	59.20	70.09	74.60	72.74	32.35	42.70	49.56

Table 6.3. Statistical characteristics of sales data in sample 2

item	1	2	3	4	5	6	7	8	9	10
min	37	24	120	124	60	68	45	78	270	239
max	335	287	413	622	245	299	350	298	938	1149
avg	107.94	105.82	223.35	266.56	125.71	140.53	133.94	169.37	542.69	565.35
s.d.	64.80	53.91	63.29	106.78	39.11	40.71	74.84	50.19	150.81	217.89

item	11	12	13	14	15	16	17	18	19	20
min	94	90	209	140	479	78	57	116	100	84
max	299	507	1134	745	2227	380	381	496	540	717
avg	183.29	201.02	499.85	341.56	895.10	179.47	149.81	242.44	219.61	247.16
s.d.	43.06	92.16	192.52	124.42	333.55	66.92	75.30	88.27	95.05	137.69

item	21	22	23	24	25	26	27	28	29	30
min	120	68	52	54	93	47	43	71	75	83
max	594	387	293	333	464	414	372	223	279	323
avg	275.10	142.61	131.65	129.35	201.03	123.19	116.27	127.74	161.67	189.23
s.d.	92.95	71.60	66.17	56.59	70.15	70.93	70.43	33.92	42.70	57.02

```
Procedure weekend-effect estimating
begin
    for gen := 1:ter-num
        input phi(gen) from KGAGA main program;
        for i :=1:mix-period-num
            total-effect(sub1) :=P-effect(sub1)*D-effect(sub1)*
                H-effect(sub1)*pre-MH-effect(sub1);
            total-effect(sub2) :=P-effect(sub2)*D-effect(sub2)*
                H-effect(sub2)*pre-MH-effect(sub2);
            n-sub1 + n-sub2 := 7;
            we-sub1 + we-sub2 := 2;
            F(i) := [lamda*total-effect(sub1)*(n-sub1 – we-sub1) +
                        lamda*total-effect(sub2)*(n-sub2 – we-sub2) +
                        lamda*total-effect(sub1)* we-sub1*phi(gen) +
                        lamda*total-effect(sub2)* we-sub2*phi(gen) ] /
                        (n-sub1- we-sub1+ n-sub2 – we-sub2 +
                          we-sub1*phi(gen) + we-sub2*phi(gen));
            error(i) := abs(sale(i) – F(i));
            total-error(gen) := total-error(gen) + error(i);
        end
    end
    for gen := 1:ter-num
        if total-error(gen) = Min(total-error)
           break;
        end
    end
    return phi(gen);
```

Figure 6.1. Pseudo codes of algorithm estimating the weekend effect via KGAGA.

Note: *sub1* denotes the first sub-period of mixed period *t*.
sub2 denotes the second sub-period of mixed period *t*.
n-sub1 denotes the length of sub1.
n-sub2 denotes the length of sub2.
we-sub1 denotes the length of the part of sub1 in the weekend.

Table 6.4. Parameters estimated via KGAGA on sample 1

item	**MAPE$_i$**	**Price Elas θ_i**	**base sales λ_i**	**P-mix $\mu_{1i(t=1:T)}$**	**P-mix $\mu_{2i(t=1:T)}$**	**P-mix $\mu_{3i(t=2:T)}$**	**P-mix $\mu_{4i(t=1:T)}$**	**P-mix $\mu_{5i(t=1:T)}$**	**P-mix $\mu_{6i(t=1:T)}$**	β_i	w_{1i}	w_{2i}	w_{3i}	w_{4i}
1	0.0468	-1.636	53.143	1.376	1.483	--	1.956	--	1.901	0.737	0.621	1.111	1.546	1.077
2	0.0392	-7.183	39.388	1.137	--	2.026	--	3.081	--	0.954	0.665	1.310	1.149	0.895
3	0.0155	-0.405	179.295	1.049	--	--	1.505	0.931	--	0.950	1.104	1.092	1.244	0.923
4	0.0212	-7.761	159.683	1.142	--	1.930	--	1.103	--	--	0.620	0.977	1.244	1.037
5	0.0223	-0.342	113.434	**0.786**	--	1.456	--	0.845	--	--	1.002	1.169	1.341	1.063
6	0.0216	-0.845	106.989	0.897	1.142	1.577	--	1.195	--	--	1.046	0.934	1.096	1.375
7	0.0352	-0.946	79.718	0.992	--	2.421	--	1.606	--	--	1.007	1.209	1.264	0.986
8	0.0260	-0.007	91.301	1.709	2.225	1.161	--	2.308	--	0.751	1.638	1.414	0.835	1.206
9	0.0183	-0.852	478.051	0.963	0.838	--	1.395	1.254	--	--	0.945	1.416	1.326	0.971
10	0.0163	-0.151	385.497	1.317	--	--	2.250	1.197	1.357	0.868	0.962	1.308	1.295	0.929
11	0.0216	-3.496	106.246	1.125	1.305	1.224	--	1.283	2.289	0.862	0.887	1.417	0.942	1.391
12	0.0246	-1.312	106.702	1.432	1.613	2.008	--	--	1.186	0.797	0.859	1.048	0.872	1.275
13	0.0189	-0.227	244.224	1.654	2.177	2.315	--	--	1.656	0.795	1.000	1.330	1.136	1.211
14	0.0222	-0.437	278.648	0.868	**0.777**	1.488	--	--	1.049	0.900	1.169	1.044	1.312	1.362
15	0.0174	-0.436	744.013	1.062	1.043	--	1.231	--	--	--	1.250	1.323	1.391	0.909
16	0.0287	-0.593	127.053	1.552	--	--	1.764	--	1.377	0.716	0.600	1.028	1.106	1.354
17	0.0279	-3.415	73.040	1.176	1.377	1.153		2.301	--	--	1.233	1.898	1.897	1.417
18	0.0227	-1.700	141.307	1.849	--	--	**1.031**	1.722	--	--	0.981	1.331	1.133	1.145
19	0.0217	-1.203	121.989	1.250	2.051	2.331	--	1.406	--	--	1.083	1.092	1.409	1.086
20	0.0255	-0.586	132.052	1.650	2.852	0.965	--	--	3.029	--	0.659	1.170	1.792	1.268
21	0.0149	0.963	221.187	**0.782**	1.234	1.636	--	--	1.647	0.919	1.293	1.445	1.052	0.989
22	0.0289	-4.634	93.058	1.001	1.196	1.723		3.007	--	--	1.003	1.526	1.060	1.080
23	0.0353	-0.181	63.999	1.160	2.932	--	2.542	2.140	--	0.975	0.606	1.575	1.190	1.896
24	0.0327	-1.511	99.059	0.840	0.977	1.663	--	--	0.930	0.897	1.250	1.268	1.306	1.052
25	0.0254	0.296	146.012	--	1.412	1.872	--	--	1.473	0.813	1.293	1.521	1.045	0.822
26	0.0394	0.003	78.669	0.903	1.293	2.976	--	--	--	--	1.168	1.140	0.727	1.385
27	0.0350	-2.247	71.403	**0.753**	1.189	2.301	--		--	--	1.220	1.614	0.729	1.135
28	0.0209	-1.839	92.552	1.594	1.463	1.330	--	1.322	1.015	0.725	1.056	0.921	0.959	1.091
29	0.0194	-0.969	122.774	1.381	1.631	1.425	--	1.472	1.523	0.775	0.911	1.256	0.823	0.984
30	0.0194	-0.325	115.196	1.488	1.915	1.422	--	1.692	2.087	0.913	0.971	1.310	0.808	0.914
mean	0.025	-1.466	162.189	1.203	1.551	1.746	1.709	1.659	1.609	0.844	1.003	1.273	1.168	1.141

Table 6.5. Parameters estimated via KGAGA on sample 2

$item_i$	***$MAPE_i$***	***Price Elas θ_i***	***base sales λi***	***P-mix $\mu_{1i(t=1:T)}$***	***P-mix $\mu_{2i(t=1:T)}$***	***P-mix $\mu_{3i(t=1:T)}$***	***P-mix $\mu_{4i(t=1:T)}$***	***P-mix $\mu_{5i(t=1:T)}$***	***P-mix $\mu_{6i(t=1:T)}$***	***β_i***	***w_{1i}***	***w_{2i}***	***w_{3i}***	***w_{4i}***
1	0.0497	-4.197	53.241	1.036	1.015	--	1.454	--	1.305	0.733	0.622	1.108	1.535	1.090
2	0.0387	-6.571	41.031	1.172	--	2.036	--	2.961	--	0.970	0.674	1.334	1.164	0.885
3	0.0164	-0.375	180.631	1.068	--	--	1.479	0.922	--	0.936	1.046	1.073	1.221	0.926
4	0.0224	-6.469	171.271	1.133	--	1.648	--	1.096	--	--	0.736	0.911	1.237	1.036
5	0.0229	-0.252	116.28	**0.778**	--	1.419	--	0.831	--	--	1.019	1.143	1.327	1.054
6	0.0225	-0.690	102.778	0.934	1.224	1.535	--	1.156	--	--	1.041	0.973	1.066	1.335
7	0.0360	-0.108	96.499	0.887	--	2.542	--	1.353	--	--	0.923	1.005	1.168	0.911
8	0.0364	-0.756	99.552	1.423	1.839	1.694	--	1.972	--	0.754	0.961	1.514	0.839	1.327
9	0.0204	-1.833	401.063	1.151	0.828	--	1.471	1.450	--	--	0.976	1.414	1.248	0.976
10	0.0177	-0.124	384.651	1.346	--	--	2.227	1.183	1.309	0.901	0.993	1.325	1.320	0.868
11	0.0226	-2.517	126.835	1.015	1.157	1.259	--	1.168	1.831	0.862	0.887	1.277	0.925	1.409
12	0.0260	-0.671	109.248	1.432	1.613	2.008	--	2.045	1.186	0.797	0.859	1.150	0.868	1.182
13	0.0217	-1.726	229.173	1.253	2.313	1.850	--	--	0.999	0.813	0.975	1.323	1.144	0.852
14	0.0229	-0.671	311.32	**0.762**	0.907	1.446	--	--	0.900	0.905	1.039	1.030	1.287	1.011
15	0.0196	-1.004	709.413	1.134	0.890	--	1.121	--	--	--	1.172	1.387	1.368	0.952
16	0.0342	-1.082	116.259	1.349	--	--	1.914	--	1.211	0.782	0.603	1.028	1.252	1.480
17	0.0300	-4.623	74.854	1.153	1.243	1.019		2.077	--	--	1.261	1.899	1.885	1.411
18	0.0235	-2.147	142.142	1.855	--	--	**0.859**	1.634	--	--	1.015	1.379	1.129	1.229
19	0.0248	-2.433	112.812	1.356	2.223	1.981	--	1.198	--	--	1.311	1.089	1.407	1.084
20	0.0308	-0.533	136.573	1.619	2.760	0.975	--	--	2.920	--	0.640	1.170	1.794	1.249
21	0.0208	-0.336	223.064	0.812	1.104	1.479	--	--	1.141	0.897	1.762	1.428	1.026	0.942
22	0.0309	-2.566	92.433	0.984	1.224	1.692	--	3.211	--	--	1.019	1.534	1.044	1.110
23	0.0385	-0.133	63.859	1.272	3.018	--	2.625	2.267	--	0.930	0.605	1.600	1.095	1.898
24	0.0381	-2.406	94.566	**0.778**	0.943	1.250	--	--	0.752	0.868	0.961	1.210	1.457	1.170
25	0.0249	-0.149	156.036	--	1.220	1.655	--	--	1.214	0.855	1.215	1.425	1.084	0.770
26	0.0431	0.436	78.387	0.951	1.162	2.812	--	--	--	--	1.066	1.173	0.748	1.390
27	0.0350	**1.492**	74.768	1.090	1.135	3.119	--	--	--	--	1.088	1.552	0.696	1.082
28	0.0250	-0.674	89.537	1.610	1.694	1.354	--	1.513	1.355	0.726	1.352	0.949	0.969	1.179
29	0.0213	-0.678	122.628	1.318	1.539	1.368	--	1.375	1.429	0.757	0.825	0.822	1.096	1.088
30	0.0224	-0.969	122.774	1.381	1.631	1.425	--	1.472	1.523	0.775	0.911	1.256	0.823	0.984
mean	0.028	-1.492	161.123	1.174	1.486	1.708	1.644	1.625	1.363	0.839	0.985	1.249	1.174	1.129

Table 6.6. Parameters estimated via OLS (enter mode) on sample 1

$item_i$	**$MAPE_i$**	**Price Elas θ_i**	**base sales λi**	**P-mix $\mu_{1i(t=1:T)}$**	**P-mix $\mu_{2i(t=1:T)}$**	**P-mix $\mu_{3i(t=1:T)}$**	**P-mix $\mu_{4i(t=1:T)}$**	**P-mix $\mu_{5i(t=1:T)}$**	**P-mix $\mu_{6i(t=1:T)}$**	β_i	w_{1i}	w_{2i}	w_{3i}	w_{4i}
1	0.049	-9.750	59.859	**0.456**	**0.407**	--	**0.786**	--	**0.535**	0.842	**0.535**	0.986	1.370	1.057
2	0.041	-7.201	42.521	1.058	--	1.927	--	2.951	--	0.887	0.647	1.212	1.141	0.890
3	0.017	-0.172	190.376	0.981	--	--	1.579	0.938	--	0.984	1.066	1.058	1.311	0.862
4	0.022	-7.647	159.812	1.117	--	1.962	--	1.168	--	--	0.611	0.976	1.269	0.988
5	0.023	0.249	113.636	0.860	--	1.669	--	0.887	--	--	0.923	1.170	1.228	1.012
6	0.022	-0.492	108.419	0.836	1.150	1.694	--	1.141	--	--	0.960	**1.922**	1.075	1.347
7	0.035	-3.009	80.238	0.990	--	1.809	--	1.392	--	--	0.916	1.209	1.245	0.971
8	0.027	**0.966**	102.000	1.726	2.140	1.346	--	2.241	--	0.820	1.441	1.692	0.874	1.068
9	0.020	0.000	502.200	0.881	**0.700**	--	1.543	1.178	--	--	0.927	1.374	1.699	1.017
10	0.017	-0.131	379.176	1.420	--	--	2.394	1.182	1.320	0.865	0.927	1.351	1.344	0.881
11	0.020	-2.502	130.713	1.005	1.134	1.155	--	1.142	1.713	0.854	0.851	1.334	0.916	1.395
12	0.026	-1.072	110.390	1.342	1.659	2.192	--	--	1.221	0.783	0.794	1.058	0.897	1.195
13	0.020	-0.038	292.656	1.523	1.761	2.151	--	--	1.505	0.754	0.935	1.377	1.074	1.068
14	0.023	**1.495**	308.895	--	1.002	2.034	--	--	1.296	--	1.024	1.085	1.361	1.265
15	0.019	0.000	712.657	1.168	1.226	--	1.837	--	--	--	1.025	1.381	1.320	0.948
16	0.028	0.105	137.689	1.697	--	--	1.879	--	1.659	0.691	**0.591**	1.081	1.126	1.249
17	0.030	-2.495	78.571	1.170	1.420	1.204	--	2.399	--	--	1.168	**1.970**	1.795	1.355
18	0.022	-2.032	141.599	1.828	--	--	0.923	1.649	--	--	0.999	1.294	1.143	1.101
19	0.031	0.000	164.350	0.983	1.553	2.387	--	1.278	--	--	0.993	1.689	1.331	1.027
20	0.026	-0.325	145.040	1.581	2.559	0.887	--	--	2.921	--	0.653	1.221	1.692	1.204
21	0.016	0.462	231.135	**0.736**	1.138	1.689	--	--	1.380	0.938	1.143	1.384	1.038	1.005
22	0.030	0.000	93.037	0.986	1.267	1.645	--	--	2.989	--	1.051	1.527	0.999	1.101
23	0.038	-0.370	69.338	1.134	2.552	--	2.303	2.201	--	0.811	**0.600**	1.377	1.109	**2.462**
24	0.034	-2.172	127.230	**0.610**	**0.681**	1.208	--	--	**0.562**	0.859	1.191	1.207	1.380	1.176
25	0.026	0.000	144.604	--	1.397	2.106	--	--	1.399	0.819	1.123	1.536	1.036	0.830
26	0.042	**1.190**	80.964	1.076	1.285	3.651	--	--	--	--	1.066	1.115	0.712	1.347
27	0.045	**1.426**	68.786	1.120	1.493	4.071	--	--	--	--	1.075	1.129	0.756	1.177
28	0.022	-2.059	92.666	1.550	1.425	1.380	--	1.251	--	0.742	1.017	0.918	0.961	1.116
29	0.021	-3.055	108.527	1.462	1.300	1.573	--	1.159	--	0.740	0.796	0.931	1.137	1.007
30	0.018	-0.796	124.210	1.241	1.629	1.330	--	1.470	1.548	0.812	0.963	1.208	0.831	0.975
mean	0.027	-1.314	170.043	1.162	1.404	1.867	1.656	1.507	1.542	0.825	0.934	1.292	1.172	1.137

Table 6.7. Parameters estimated via OLS (enter mode) on sample 2

item$_i$	**MAPE$_i$**	**Price Elas θ_i**	**base sales λi**	**P-mix $\mu_{1i(t=1:T)}$**	**P-mix $\mu_{2i(t=1:T)}$**	**P-mix $\mu_{3i(t=1:T)}$**	**P-mix $\mu_{4i(t=1:T)}$**	**P-mix $\mu_{5i(t=1:T)}$**	**P-mix $\mu_{6i(t=1:T)}$**	β_i	w_{1i}	w_{2i}	w_{3i}	w_{4i}
1	0.056	-4.171	60.886	0.897	0.908	--	1.402	--	1.154	0.766	0.760	0.969	1.347	1.057
2	0.040	-7.130	42.777	1.059	--	2.052	--	2.930	--	0.888	0.733	1.212	1.143	0.889
3	0.017	-0.172	190.380	0.981	--	--	1.579	0.938	--	0.984	1.066	1.058	1.311	0.862
4	0.023	-6.804	166.002	1.117	--	1.868	--	1.168	--	--	0.696	0.939	1.269	0.988
5	0.024	0.249	114.550	0.853	--	1.587	--	0.880	--	--	1.062	1.161	1.228	1.012
6	0.023	-0.307	110.280	0.821	1.149	1.644	--	1.138	--	--	0.936	0.907	1.058	1.326
7	0.036	-1.270	85.285	0.942	--	2.435	--	1.422	--	--	0.957	1.137	1.245	0.971
8	0.040	-0.926	113.300	1.420	1.586	1.278	--	1.813	--	0.750	1.473	**1.852**	0.836	1.168
9	0.021	-1.786	459.440	0.963	**0.565**	--	1.214	1.288	--	--	1.126	1.374	1.699	1.017
10	0.018	-0.439	382.220	1.338	--	--	2.109	1.125	1.240	0.856	1.138	1.406	1.322	0.874
11	0.023	-2.502	122.610	1.071	1.209	1.338	--	1.218	1.828	0.854	0.956	1.334	0.916	1.395
12	0.027	-1.072	110.390	1.342	1.659	2.401	--	2.314	1.221	0.783	0.939	1.058	0.898	1.195
13	0.023	-0.536	284.580	1.416	1.811	2.223	--	--	1.281	0.747	1.057	1.377	1.064	0.954
14	0.025	-3.454	308.900	**0.493**	**0.605**	--	--	--	**0.629**	0.922	1.023	1.085	1.361	1.012
15	0.021	-0.405	686.080	1.212	1.090	--	1.627	--	--	1.000	1.096	1.433	1.320	0.985
16	0.034	-0.356	123.220	1.677	--	--	1.990	--	1.627	0.689	0.700	1.110	1.125	1.397
17	0.032	-5.126	78.571	1.188	1.197	0.995	--	2.020	--	--	1.294	**1.941**	**1.795**	1.355
18	0.024	-2.253	141.460	1.831	--	--	0.898	1.605	--	--	1.001	1.274	1.143	1.074
19	0.037	-2.347	154.780	1.044	1.649	1.584	--	0.859	--	--	1.103	1.689	1.331	1.027
20	0.032	-0.325	140.050	1.638	2.651	0.980	--	--	3.025	--	0.896	1.221	1.692	1.204
21	0.022	-1.909	219.862	0.835	1.026	1.357	--	--	**0.760**	0.868	1.379	1.455	0.962	0.931
22	0.031	-2.561	93.037	0.986	1.267	1.745		2.989	--	--	1.046	1.527	1.000	1.101
23	0.041	-0.266	70.739	1.122	2.565	--	2.545	2.192		0.809	0.745	1.387	1.110	**2.450**
24	0.039	-2.172	103.440	**0.750**	0.855	1.179	--	--	**0.691**	0.859	1.169	1.207	1.380	1.176
25	0.026	**2.253**	145.619	--	1.618	2.588	--	--	2.545	0.859	1.169	1.525	1.103	0.824
26	0.043	**2.045**	85.115	1.093	1.214	3.209	--	--	--	--	1.192	1.123	0.717	1.281
27	0.049	**1.071**	71.665	1.046	1.433	2.858	--	--	--	--	1.289	1.129	0.725	1.130
28	0.025	-0.747	94.066	1.525	1.613	1.318	--	1.504	1.313	0.742	1.297	0.903	0.961	1.116
29	0.022	-0.624	113.069	1.404	1.616	1.480	--	1.613	1.637	0.740	0.938	0.893	1.137	1.007
30	0.023	-0.796	120.904	1.275	1.672	1.513	--	1.510	1.590	0.812	1.099	1.241	0.831	0.975
mean	0.030	-1.495	166.443	1.150	1.407	1.792	1.670	1.607	1.467	0.829	1.044	1.264	1.168	1.125

Table 6.8. Collinearity diagnosis across items via Maximal VIF for OLS47 and OLS56 in regression

item	1	2	3	4	5	6	7	8	9	10
VIF-OLS47	**114.96**	3.355	2.160	2.351	2.152	2.881	**12.188**	4.658	2.665	**7.289**
VIF-OLS56	**6.066**	3.440	2.073	2.667	2.603	3.279	**5.707**	3.243	3.579	3.355

item	11	12	13	14	15	16	17	18	19	20
VIF-OLS47	4.464	**9.621**	**8.190**	1.452	1.231	2.834	**5.277**	**7.243**	1.825	2.170
VIF-OLS56	3.018	**10.623**	**7.997**	1.926	**5181.792**	3.184	3.526	**10.718**	3.073	2.072

item	21	22	23	24	25	26	27	28	29	30
VIF-OLS47	**101.457**	1.557	1.819	**8.662**	2.422	3.709	3.617	2.326	2.326	**5.786**
VIF-OLS56	**18.710**	1.750	1.895	**7.614**	**40.298**	3.876	4.386	**7.712**	**7.712**	**6.479**

6.3.3 Comparing and analyzing results from various kinds of forecast adjustment methods

The performance of weekly sales forecasting adjustment from various methods in terms of MAPE can be displayed in Table 6.9 and Table 6.10. Each cell with negative adjustment performance is in boldface. Among these adjustment methods, without taking advantage of any adjustment, the MAPE of sales forecasting with the regression model, that is NA, in average, is 25.75% and 32.27% in busy season and off season, respectively.

If forecasts are adjusted with SIR (seasonal index realignment), the average performance across items in terms of MAPE is 26.49% for busy season, which seems a little worse than NA, check Table 6.9. However, for off season, the average figure of SIR is 29.58%, an improvement of mere 8.34% over initial forecast, check Table 6.10. There are 16 items improved out of 30 because of SIR for busy season, while there are 17 items get improved due to SIR's contribution for off season. The relatively poor performance of SIR in busy season may be attributed to its instability in improving the original forecasts and the relatively good performance of NA compared to that of ARIMA without adjustment. Overall impact of SIR in both seasons seems to be trivial, check Table 6.9 and Table 6.10. However, if combined with PA (results in TA), it does make a big difference.

If adjustment is conducted with PA (proportional adjustment of mixed effect in mixed periods and level adjustment) in busy season, we see an improvement from 25.75% to 23.84%, a 7.42% improvement in average, it still is not very impressive. The number of items with negative results is 10, there are 13 adjustments out of 24 in total get improvement in MAPE in busy season, check Table 6.9. However, for off season, it's a different story for PA, in off season, there are 21 adjustments out of 25 in total get the forecasts improved. Average MAPE is reduced from 32.27% to 23.81% by PA, a remarkable 26.22% improvement.

If TA (total adjustment) is performed, since it combines the adjustments from both SIR and PA, it offers the performance better than that of all other adjustments in this study, for busy season, average MAPE reduced from 25.75% to 22.28%, the improvement of MAPE amounts to 13.48% in average. In particular, in off season, the improvement is even more impressive, the average MAPE reduced from 32.27% to 21.63%, a very impressive 32.97% improvement over MAPE of initial forecasts.

Table 6.9. Comparison of the accuracy of various forecast adjustment methods in busy season 2008

MAPE of various forecast adjustment methods						
item	**NA**	**PA**	**SIR**	**TA**	**ARIMA**	**ES**
1	0.2428	**0.3537**	0.2327	0.1811	0.4501	0.4078
2	0.2330	**0.3270**	0.1468	0.0998	0.3097	0.3656
3	0.0787	0.0787	**0.1468**	**0.1468**	0.2334	0.1678
4	0.1099	0.1099	**0.2477**	0.1099	0.1608	0.1512
5	0.0730	**0.1050**	**0.2080**	**0.1850**	0.1425	0.1601
6	0.2150	0.2130	0.2120	0.2120	0.1429	0.1573
7	0.2370	0.2330	0.1420	0.1840	0.2754	0.3198
8	0.5143	0.5143	0.4910	0.4910	0.2374	0.1890
9	0.1671	**0.198**0	**0.2889**	**0.2638**	0.2001	0.1890
10	0.2180	0.2120	**0.2407**	0.1850	0.2573	0.2964
11	0.2347	0.1258	0.1764	0.0468	0.1213	0.1081
12	0.2400	0.2400	0.1676	0.1348	0.3810	0.3664
13	0.2464	0.2279	0.2016	0.1874	0.3926	0.2563
14	0.2165	**0.2639**	0.1471	0.1689	0.2619	0.1858
15	0.2469	**0.3235**	0.2192	0.2229	0.1426	0.1113
16	0.1586	**0.1609**	**0.2204**	**0.1667**	0.2108	0.2379
17	0.1548	**0.1580**	**0.1604**	**0.1582**	0.1964	0.1823
18	0.1379	**0.1462**	**0.2372**	**0.2297**	0.2119	0.7748
19	0.3311	0.1697	**0.5476**	**0.3799**	0.2154	0.2309
20	0.2733	0.2505	**0.2840**	0.2625	0.1457	0.1789
21	0.2770	0.2770	0.1213	0.1463	0.2128	0.1710
22	0.3457	**0.3585**	0.1761	0.1866	1.1039	1.7805
23	0.3172	0.3172	0.2671	0.1914	0.2479	0.2398
24	0.6743	0.4641	**0.9168**	0.6597	0.4356	0.5149
25	0.2428	0.2428	**0.3637**	**0.3251**	0.2304	0.2405
26	0.5445	0.3226	0.2948	0.3068	0.1842	0.2711
27	0.4078	0.2554	**0.5561**	0.3744	0.1410	0.1997
28	0.2427	0.2290	0.1570	0.1448	0.2558	0.2542
29	0.0904	0.0783	**0.1637**	**0.1331**	0.2431	0.1407
30	0.2526	0.1975	0.2128	0.1989	0.2559	0.2388
avg	0.2575	0.2384	0.2649	0.2228	0.2960	0.3230
std	0.1350	0.1041	0.1669	0.1238	0.1800	0.3100

Table 6.10. Comparison of the accuracy of various forecast adjustment methods in off season 2008

MAPE of various forecast adjustment methods						
item	**NA**	**PA**	**SIR**	**TA**	**ARIMA**	**ES**
1	0.2926	0.2584	0.2679	0.2337	0.2716	0.3561
2	0.1744	0.0942	**0.2927**	**0.2082**	0.2051	0.5534
3	0.1948	0.1948	0.1445	0.1445	0.0931	0.0859
4	0.1980	**0.2000**	**0.2460**	0.1920	0.0931	0.1364
5	0.2465	0.2465	0.1888	0.1888	0.1551	0.2050
6	0.1170	**0.1240**	0.0931	0.0947	0.2409	0.0838
7	0.2271	0.2271	0.2192	0.2192	0.2601	0.1224
8	0.3213	0.2774	**0.3639**	0.3200	0.3667	0.2386
9	0.6842	0.5906	0.4551	0.5010	0.2028	0.2215
10	0.3295	0.1534	**0.3698**	0.2199	0.0905	0.1701
11	0.8858	0.6513	0.7132	0.5005	0.3382	0.5336
12	0.1148	0.1148	0.1000	0.1000	0.6658	0.1502
13	0.5338	0.5085	0.1926	0.2058	0.1009	0.1081
14	0.1837	0.1837	0.1801	0.1801	0.7285	0.4012
15	0.7549	0.2158	0.7310	0.1857	0.2671	0.1507
16	0.5693	0.1591	0.4662	0.1098	0.2958	0.1563
17	0.2388	0.2224	**0.3385**	**0.3137**	0.4142	0.2163
18	0.3875	0.3218	0.3498	0.2841	0.2933	0.3826
19	0.1258	0.0910	**0.1956**	**0.1595**	0.3125	0.4823
20	0.2619	0.1881	**0.2937**	0.1924	0.6223	0.4630
21	0.2161	0.1927	0.2161	0.1926	0.1530	0.5128
22	0.4593	0.3087	**0.4687**	0.3245	0.1375	0.1508
23	0.2521	0.1958	0.1189	0.1301	0.2757	0.2785
24	0.1377	0.1377	0.0789	0.0789	0.4395	0.3717
25	0.3147	0.2577	0.2090	0.1403	0.4383	0.1815
26	0.3106	0.3084	**0.3461**	**0.3171**	0.2958	0.2613
27	0.2882	0.2069	**0.2940**	0.2062	0.2521	0.2658
28	0.3765	0.1728	0.2793	0.0904	0.2381	0.2381
29	0.2002	**0.2242**	**0.2757**	**0.2935**	0.2183	0.2045
30	0.2850	0.1150	**0.3851**	0.1621	0.3026	0.2743
avg	0.3227	0.2381		0.2163	0.2920	0.2650
std	0.1910	0.1333	0.1579	0.1048	0.1610	0.1390

If the adjustment of ARIMA is of concern, 20 out of 30 items get improved in busy season, check Table 6.9, but no sign of MAPE improvement in average is recorded. The effectiveness of this adjustment doesn't look as good as that of PA and TA, as well as SIR.

In off season, the adjustment of ARIMA displays a performance obviously better than its performance in busy season, the number of items get improved in MAPE reduced to 18, a bit worse than that in busy season. However, overall percentage of improvement from adjustment amounts to 13.78%, a performance much better than its counterpart in busy season.

As adjustments of ES are of concern, its performance in improvement of average MAPE is about the same with that of ARIMA, just the number of improvements in total adjustment reduced to 18 out of 30 in total in busy season. In off season, the adjustments of ES show a negative improvement of initial forecasts in average, its MAPE increased from 25.50% to 26.52%, a -4% improvement in average. Besides, the number of improvements in adjustment significantly reduced to 12 out of 30 adjustments in total. Compared with ARIMA in off season, the average MAPE of ES after adjustment still is better than that of its counterpart by 9.27% in average difference.

In the Appendix, Figure A1-A30 show the detail of sales data in the training period of 2007 and 2008, Figure B1-B30 portray multi-step out-of-sample forecasts of various forecasting methods and adjustment methods in detail in the busy season of 2008, Figure C1-C30 shows in detail multi-step out-of-sample forecasts of various forecasting methods and adjustment methods in the off season in 2008. In most cases, points of forecasts made by TA adjustments or PA adjustments change somewhat in accordance with the promotion activities and moving holidays. However, points of forecasts made by ARIMA, ARIMA's adjustments and ES and ES' adjustments, in almost every case, usually form a linear shape and don't cope with the dynamic changes of original unit sales even in the least sense, showing that they are limited to mainly stationary sales forecasting and are insensitive to turning points of sales of promotions and holidays, owing to the fact that they are mostly blind to contextual information in this study.

6.4 Analysis of Various Adjustment Methods from the Perspective of Adjustment Size

In this subsection, the percentage of correct direction in adjustments over total adjustments is used to measure the performance of various adjustment methods in a variety of adjustment sizes. Whether the direction is correct or not depends on the comparison between initial forecast and the actual observation, if initial forecast is under-forecast, the correct direction of adjustment should be adjusted upwards, regardless of adjustment size. On the other hand, if initial forecast is over-forecast, the correct direction of adjustment should be adjusted downwards, regardless of adjustment size. However, if the initial forecast is within the range of [actual unit sales - 3%*actual unit sales, actual unit sales + 3%*actual unit sales], any subsequent adjustment with result less than or equal to initial over-forecast or any adjustment with result more than or equal to the under-forecast is considered to be adjustment in the correct direction.

In this study, any adjustment with result within less than 10% range of the initial forecast, regardless of adjustment direction, is regarded as small adjustment, otherwise, it's a large adjustment. In Table 6.11, all the small adjustments, in both busy season and off season, have the ratio of adjustment with correct direction, regardless of adjustment method, to be less than or at most equal to 60%, except that of PA in busy season which is 66.67%.

On the other hand, large size adjustments seem to have a much more consistent and better performance in average than that of small ones, with the average correct-direction ratio at least over 57% in off season for the proposed adjustment methods in this study.

As for that of ARIMA and ES, both methods only have large adjustments (± 20%), show a correct-direction ratio under 50%, a sharp contrast with their counterpart in busy season which demonstrate a ratio higher than 81% indeed. In busy season, an average ratio of correct direction above 81% is recorded for methods proposed in this study, except that of SIR in busy season which only has 67.83%. This consequence is not surprising, in the literature, there are considerable similar evidences (Fildes and Goodwin, 2007; Syntetos et al., 2009).

As Michael Lawrence et al. (2006) pointed out, the adjustments for domain knowledge were overall beneficial, but were most advantageous when large adjustments were made. When small adjustments were made, they seemed to be less than useful, perhaps reflecting the tendency to tinker at the edges.

Among three adjustment methods proposed in this thesis, PA seems to have the most robust performance in terms of the ratio of correct-direction adjustment in both small and large adjustments in busy season, while SIR and TA in the same season have a poor performance with a correct-direction ratio less than 50% in small adjustments.

However, in off season with small adjustments, PA seems to be the loser, with an overall ratio of correct direction adjustment less than 50%, whereas with large adjustments in off season, PA still is the best performer in terms of the ratio of correct-direction adjustment. Whether this adjustment provides the most positive contribution to forecast accuracy improvement or not, it seems that we still have to crosscheck with other criteria like IMP in Table 6.12 and MAPE in Table 6.10 to have an adequate assessment. Since an adjustment made in the correct direction is just one of the fundamental prerequisite for an adjustment to make an improvement, another important requirement is the magnitude of an adjustment must not be too large particularly for large size adjustments, otherwise it is very easy to have over-adjustment resulting in negative improvement.

A measure called IMP, which can be used to evaluate the adjustment improvement, may be formulated below:

$$IMP = APE_{ini} - APE_{ad} \tag{6.5}$$

Where, APE denotes absolute percentage error, APE_{ini} stands for APE of initial forecast, while APE_{ad} stands for APE after adjustment.

In Table 6.12, with the only exception of SIR applied in off season, large adjustments consistently and significantly outperform small adjustments in terms of IMP, regardless of the adjustment method. The only exception of SIR implies that a lot of large size SIR adjustments with correct direction in off season in Table 6.11 are actually over-adjusted, its number is much more than that of small size SIR adjustments in the same season. Note that in busy season, all three adjustment methods using small adjustment, the average IMP are all negative except that of PA, among them, more than half of small adjustments made by SIR and TA are in correct direction, this means that in average the contribution made by these adjustments are canceled by adjustments in wrong direction.

Table 6.11. Comparing the performance concerning direction of adjustment of various adjusting methods

	busy season 2008				off season 2008			
AD method	ratio of small ad	% of correct direction in small ad	ratio of large ad	% of correct direction in large ad	ratio of small ad	% of correct direction in small ad	ratio of large ad	% of correct direction in large ad
PA	24/76	66.67%	52/76	82.69%	24/88	**45.83%**	64/88	92.19%
SIR	37/153	37.84%	115/153	68.10%	60/180	**58.33%**	120/180	59.17%
TA	40/159	47.50%	119/159	83.33%	50/180	64.00%	130/180	68.46%
ARIMA	0/180	--	180/180	82.22%	0/180	--	180/180	41.67%
ES	0/180	--	180/180	81.11%	0/180	--	180/180	45.56%

Table 6.12. Comparing IMP of various forecast adjustment methods in either small or large adjustments

	busy season 2008		off season 2008	
AD method	avg IMP from small adjustments	avg IMP from large adjustments	avg IMP from small adjustments	avg IMP from large adjustments
PA	2.29%	8.28%	**3.13%**	24.50%
SIR	-1.36%	-0.54%	**0.028%**	-3.27%
TA	-0.10%	7.71%	4.32%	9.72%
ARIMA	--	3.67%	--	-0.75%
ES	--	2.78%	--	-3.79%

In busy season, PA, which encompasses proportional adjustments related to holidays and promotions and adjustments of level change, is the best performer in terms of IMP, it also is the best one even from the viewpoint of ratio of correct-direction adjustment, crosscheck with Table 6.9, obviously, it provides the most consistent and the largest contribution to the improvement of forecast accuracy per adjustment in average in busy season among various methods. In Table 6.13, the ratio of positive IMP from small adjustments and large adjustments is a whopping 70.83% and 73.08%, way better than its counterparts of SIR and TA in their 40s percent and at most 60%, respectively.

However, in busy season, because of the relatively less frequency of PA adjustments made, check Table 6.11, even though it offers the best performance in terms of IMP and ratio of positive IMP, check Table 6.12 and Table 6.13, its overall contribution to improvement of forecast accuracy in terms of MAPE in busy season is not very impressive, check Table 6.9.

As large size adjustments are concerned, in Table 6.7, even though PA is not the best performer in terms of the ratio of correct-direction adjustment in busy season, however, in Table 6.12, it does have the best performance in terms of IMP, this implies that other factors may determine their contributions as well.

In addition, if we check both Table 6.11 and Table 6.12, there are other issue of inconsistency between these two tables, for instance, in off season, the ratio of correct-direction adjustments in small adjustments for PA in Table 6.11 is a mere 48%, however, in Table 6.12, PA has a good performance of 3.13% even beats SIR in average IMP from small adjustments which has a relatively much better ratio of correct-direction at 55.93%.

Hence, in Table 6.13, a new measure called ratio of positive IMP (of all adjustments) is used to assess the performance of various methods of adjustment, the reason is that only adjustments with positive IMP actually improve forecast, because it exclude adjustments of over-adjustment in correct-direction adjustments.

On the other hand, although the performance of TA in terms of average IMP and ratio of positive IMP in both seasons is not the best among these methods demonstrated, due to its highest number of adjustment made, check Table 6.11, TA still provides the most positive contribution to the improvement of forecast accuracy in terms of MAPE, check Table 6.9 and Table 6.10, because it provides the most holistic adjustment among three methods.

The contribution of adjustments of SIR, which actually is seasonal index realignment, though is not obvious or even mostly negative in terms of average IMP in Table 6.12, if combined with the adjustment of PA, that is TA, provides the most contribution to forecasting accuracy as mentioned above.

Table 6.13. Further analysis of various forecast adjustment methods in adjustments on ratio of positive IMP

	busy season 2008		off season 2008	
AD method	Ratio of positive IMP from small adjustments	Ratio of positive IMP from large adjustments	Ratio of positive IMP from small adjustments	Ratio of positive IMP from large adjustments
PA	70.83%	73.08%	62.50%	87.50%
SIR	40.54%	44.83%	**51.67%**	48.33%
TA	45.00%	60.00%	60.00%	60.00%
ARIMA	--	73.89%	--	34.44%
ES	--	73.33%	--	31.67%

As for adjustments of ARIMA and ES, with fixed adjustment size at 20%, both works quite well in terms of the ratio of positive IMP of all adjustments in busy season, check Table 6.13, but their average IMP of adjustments are all less than 4%, all are less than half of those of PA and TA in the same season. This is because in case of right-direction adjustment the improvement in IMP usually is not big, however, if an adjustment is wrong-direction, with 20 % in adjustment size, the negative improvement of accuracy usually is less than -20%, even more than 80% of adjustments from ARIMA and ES are correct-direction, most of their contribution are eroded by wrong-direction adjustments and over-adjustments in correct-direction adjustments. In off season, the performance of ARIMA and ES adjustments is even worse, in terms of average IMP, they are all negative, in terms of the ratio adjustments with positive IMP in all adjustments are just mere 34.44% or even less.

6.5 Analysis of Various Adjustment Methods from the Perspective of Lead Time

In last subsection, we analyzed the performance of various adjustment methods from the viewpoint of adjustment size in terms of the ratio of correct-direction and IMP as well as the ratio of positive IMP. In this subsection, we will analyze them in terms of IMP from the perspective of lead time.

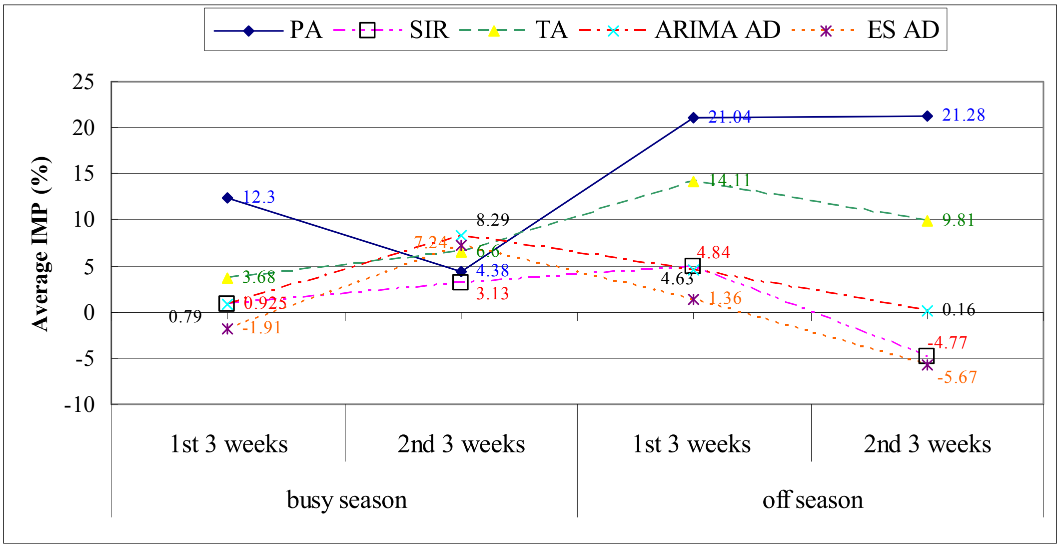

Figure 6.2. Comparison of average IMP of various adjustment methods on different lead times

In Figure 6. 2, the forecasting horizon is divided into two parts, namely, the first 3 weeks and the second 3 weeks in both busy season and off season. Obviously, in busy season, the performance of various adjustment methods is relatively more stable than that of its counterpart in off season in which the spectrum of IMP is more spread out than that in the busy season.

Among different adjustment methods, PA seems to have the best performance in terms of IMP across different seasons, its average IMP in off season even touches 21% in both periods, but in the second period of busy season, its performance turned downwards. TA ranked second, and SIR still is the worst performer which shows an average IMP of -4.77% in the second period of off season.

The performance of adjustment of ARIMA looks not so stable in different lead time of busy season, however, in the second period it shows an average IMP up to 8.29%, in off season, its average IMP still stays above zero. While the average IMP of ES adjustment in the first period in busy season and the second period in off season all appear negative.

In general, in the busy season, except the adjustment of PA, the performance of various adjustments doesn't go down as lead time increases, as a matter of fact, most of them even go upward in the second 3 weeks. In contrast, the performance of adjustments in the off season is quite different, average IMP seems decreasing as lead time increases from the first 3 weeks to the second weeks with the only exception of PA adjustment which still stays unchanged.

6.6 Summary

From the above analysis and explanations, on a per adjustment basis, PA offers the most effective accuracy improvement of forecasts in 3 out of 4 measures in terms of both average IMP and ratio of positive IMP in both seasons, check Table 6.12 - 6.13, it also has the highest percentage in correct-direction adjustments in busy season with small adjustment and large adjustments made in off season as well. Since it is relatively less frequently used, check Table 6.11 (the mixed effect condition and the level change don't arise as frequently as seasonal index realignment in general), its own total contribution is not very impressive.

TA, on the other hand, a combination of PA and SIR, is more considerate in applying contextual information to reflect changes in promotions and holidays in both seasons, and provides the most contribution in improving MAPE, even though in terms of percentage of correct-direction adjustments, average IMP and the ratio of positive IMP, it rarely is the best performer. Crosscheck Table 6.11-6.13, it is easy to see that, correct direction is the prerequisite for an adjustment of any kind to improve forecasting accuracy, but due to the issue of over-adjustment, many adjustments of correct direction still make negative contributions to forecasting accuracy. In contrast, the ratio of positive IMP is a more robust measure to assess the performance of various adjustment methods mentioned above due to its exclusion of over-adjustment in measuring the performance of improvement from adjustments.

As adjustment of ARIMA is concerned, its performance in terms of average IMP and the ratio of positive IMP is not as good as three methods proposed in this study, its fixed-size adjustment still provides positive contribution to improve the forecasting accuracy of initial forecasts in both seasons. While adjustments of ES mostly don't look very promising except those in busy season.

Besides, with the only exception of SIR in off season, if the performance is measured in terms of both percentage of correct-direction adjustment, average IMP, and the ratio of positive IMP in average, large size adjustment, in general, has a significant advantage over small size adjustment, regardless of season.

Our model assumes that there will be no big difference between actual promotion activities and those specified in promotion proposals in forecast horizon, if this is not true, there will be larger MAPE incurred for the original forecasts and various types of forecast adjustments as well.

The relatively more accurate performance of original forecasts in busy season than in off season may due to the fact that promotion and calendar effects are so strong that they dominate unit sales in busy season, while in off season, these effects are much less obvious and much less frequent as in busy season, other factors like seasonal index realignment, competitors' actions and so on may have critical impacts on unit sales therein.

7. Conclusions

If the sample size is large enough or there isn't any high correlation among critical predictor variables, multiple linear regression with typical least square estimator like ordinary least square can be an optimal tool in analysis and prediction of time series data, because large variation of value in observation data related to certain critical predictor variable is sufficient to support an adequate regression modeling. As a result, relevant domain knowledge or contextual information can be put into a good use.

However, in case the regression modeling is based on limited size sample, or for whatever reason, the variation of values in the dataset pertain to specific variable of the model is not enough, the issue of collinearity or multi-collinearity will arise, in such a way that it would negatively and seriously affect proper variable identification and variable coefficient estimation. Under such situation, any analysis or forecasting based on these model parameters thereafter can be questionable.

We propose an alternative estimator—a knowledge guided adaptive genetic algorithm (KGAGA) with proper formulation in fitness function and realistic constraints of coefficient of critical variables. In particular, a detect and escape mutation algorithm (DEMA) via a MAFI(l) as measure of local pitfall and feedback information after action taken through the employment of three (at most) different types of mutation operators in different phases to keep population diversity and convergence ability at the same time to help the search come out of the local trap (in a broad sense) many times and, eventually, the probability of convergence to a optimal or near optimal solution is significantly enlarged. In such a way the search capability of ordinary KGAGA is considerably improved, such that KGAGA has a better and more consistent performance in parameter estimation.

The forecasting adjustment mechanism proposed in this thesis concerns with realignment of seasonal indices and the anticipated mixed effect of certain variables, such as the multiplier of the effect of promotion mix, and the multiplier of holiday effect, already incorporated in the regression model and assessed with KGAGA which is more flexible and is capable of better deriving realistic coefficient of variables from limited sizes sample than most other conventional alternatives.

The mechanical adjustment mechanism proposed in this thesis becomes a necessary and natural extension of the regression model which doesn't take variations of mixed effect of promotions and holidays in forecast horizon from those in training period into account and thus incapable of handling them. And in the process of forecast adjustment, subjective judgment based on contextual information is minimized.

Among three adjustment methods embedded in the adjustment mechanism of this thesis, PA, providing the necessary reassessment of mixed effect in mixed periods and level adjustment, is capable of offering the most contribution to the improvement of forecasting accuracy on per adjustment basis. However, the frequency of variation of mixed effect related to promotions and holidays as well as level changes is not high in both seasons, causing total effect in improving accuracy of forecasts in terms of MAPE become very limited. On the other hand, SIR focuses on realignment of seasonal index in forecasting horizon in a different year based on seasonal index of referenced periods, provides very limited effect in improving original forecast alone in terms of IMP and MAPE. Surprisingly, combine PA and SIR to form TA showing the most effective consequence in improving MAPE of original forecasts, even though TA doesn't look as promising as PA in various measures on per forecast basis at both seasons. This is due to the fact that it is the most comprehensive adjustment method in coping properly with anticipated variations of sales in the forecasting horizon before hand.

Adjustments of ARIMA and ES with fixed percent in both seasons seem too rough and unable to reflect the dramatic change of sales at the turning point of variation of promotions and holidays, this can be demonstrated in charts plotted with forecasts and related adjustments in a direct comparison with those by PA and TA and so on, and the importance of incorporating contextual information or domain knowledge into the parameter estimator and adjustment mechanism of the forecasting system is thus confirmed.

References

1. Abuizam, R., Thomopoulos, N. T. (2005). Adjusting an existing forecasting model when some future demands are known in advance; A bayesian technique. Working Paper. Illinois Institute of Technology, Chicago, Illinois.
2. Adnan, N., Ahmad, M.H., and Adnan, R. (2006). Comparative study on some methods for handling multi-collinearity problems, *Matematika,* 22, 109-119.
3. Allenby, G., Ginter, J. L. (1995). The effect of in-store displays and feature advertising on consideration sets, *International Journal of Research in Marketing,* 12, 76-80.
4. Armstrong, S. (1984). Forecasting by extrapolation: Conclusion from 25 years of research. *Interface*s 14, 6, 52--66.
5. Armstrong, S. (1985). *Long-range forecasting: from crystal ball to computer*, New York; Wiley.
6. Armstrong, J. (2006) Findings from evidence-based forecasting: Methods for reducing forecast error, *International Journal of Forecasting*, 22 (3), 583-598.
7. Armstrong, J.S., and Collopy, F. (1992) Error measures for generating about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8, 69-80.
8. Armstrong, J.S., and Collopy, F. (1998) Integration of statistical methods and judgment for time series forecasting: principles from empirical research, in G. Wright and P. Goodwin (eds.), Forecasting with judgment. Chichester, England: John Willey, 263-293.
9. Armstrong, S., Collopy, F., & Thomas, Y. J. (2005). Decomposition by causal forces: a procedure for forecasting complex time series. *International Journal of Forecasting*, 21, 25–36.
10. Belsley, D. A. (1982). Assessing the presence of harmful collinearity and other forms of weak data through a test for signal-to-noise. *Journal of Econometrics*, *20*, 211--253.
11. Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity.* New York: John Wiley & Sons.
12. Berilant, J., de wet, T., & Goegebeur, Y. (2005). A goodness-of-fit statistic for Pareto-type behavior. *Journal of Computational and Applied Mathematics*, 186, *1*, 99-116.
13. Blattberg. R. C. and Hoch, S. J. (1990). Database models and managerial intuition ! 50% model + 50% manager. *Management Science*. 36. 8, 887--899.
14. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). Englewood Cliffs, NJ7 Prentice Hall.
15. Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J., Kell, D. B., (1997), Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry, *Analytica Chimica Acta*, 348, 71-86.
16. Bunn, D. W., & Salo, A. A. (1996). Adjustment of forecasts with model consistent expectations. *International Journal of Forecasting*, 12, 163—170.
17. Bunn, D. and Wright, G. (1991) Interaction of judgmental and statistical forecasting

methods: Issues and analysis. *Management Science* 37, 501-518.
18. Burger, C.J.S.C., Dohnal, M., Kathrada, M., Law, R., (2001) A practitioners guide to time-series methods for tourism demand forecasting –a case study of Durban, South Africa. Tourism Management 22 (4), 403-409.
19. Burt, R. S. (2000). Decay functions, *Social Networks*, 22, 1-28.
20. Buser, O. (1983). Avalanche forecast with the method of nearest neighbors: An interactive approach. *Cold Regions Science Technology*, 8, 115–163.
21. Camerer, C. F. (1981). General Conditions for the success of Bootstrapping Models, *Organizational behavior and Human Performances*, 27, 411-422.
22. Carroll, R.J., and Ruppert, D. (1988) *Transformation and Weighting in regression*. Chapman and Hall. New York. NY. USA.
23. Chen, Y.,. Hu, J., Hirasawa, K., and Yu, S. (1997). GARS: An Improved Genetic Algorithm with Reserve Selection for Global Optimization, in Proceedings of the 9th annual conference on Genetic and evolutionary computation, 1173-1178.
24. Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836.
25. Coenders, G., and Saez, M. (2000). Collinearity, heteroscedasticity and outlier diagnostic in regression. Do they always offer what they claim? *New Approaches in Applied Statistics*, 16, 79-94.
26. Cook, R. D. (1977) Detection of influential observations in linear regression. *Technometrics* 19, 15-18.
27. Cooper, L. G., Baron, P., Levy, W., Swisher, M., & Gogos P.(1999). PromoCast?: A New Forecasting Method for Promotion Planning. *Marketing Science*, 18(3), 301-316.
28. Cooper, L. G., & Giuffrida, G. (2000). Turning Datamining Into a Management Science Tool. *Management Science*, 46(2), 249-264.
29. Craney, T.A., and Surles, J. G. (2002) Model-dependent variance inflation factor cutoff values. *Quality Engineering* 14(3), 391-403.
30. Dalrymple, D. J. (1987). Sales forecasting practices: Results from a U. S. survey, *International Journal of Forecasting*, 3, 379-391.
31. De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22, 443– 473.
32. De Jong, K. A. and Spears, W. M. (1992). A formal analysis of the role of multi-point crossover in genetic algorithms. *Annals of Mathematics and Artificial Intelligence* 5(1), 1–26.
33. Donselaar, K. van., Kopczak, L. R., Wouters, M. (2001). The Use of Advance Demand Information in a Project-Based Supply Chain, *European Journal of Operations Research*130 (2001) 519--538.
34. Draper, N. R., & Smith, H. (1998). *Applied regression Analysis,* John Wiley & Sons, INC. New York, NY, USA.
35. Duvvuri, Sri Devi., Ansari, Asim., and Gupta, Sunil. (2007). Consumers' Price Sensitivity Across Complementary Categories, *Management Science*, 53 (12), 1933-1945.

35. Edmundson, R. H., (1990). Decomposition: A strategy for judgmental forecasting. *Journal of Forecasting*, 19, 305—314. *IEEE Transactions on Evolutionary Computation*, 3(2), 124–141.
36. Edmundson, R., Lawrence, M., & OConnor, M. (1988). The use of non time series information in time series forecasting. *Journal of Forecasting*, 7, 201—212.
37. Edmundson, R.H. (1990). Decomposition: A strategy for judgmental forecasting. *Journal of Forecasting* 9, 305-314
38. Eiben, A.E., Hinterding, R., and Michalewicz, Z. (1999) Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation* 3(2), 124-141.
39. Eye, A.V. (2006). Comparing Tests of Multinormality under Sparse Data Conditions - a Monte Carlo Study. interstat.statjournals.net
40. Fildes, R., & Goodwin,P. (2007). Good and bad judgement in forecasting: Lessons from four companies. *Foresight*, Fall, 5—10.
41. Foekens, E. W., Leeßang, Peter S. H., & Wittink, D. R. (1999). Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometric*s, 89, 249–268.
42. Franses, P. H., & McAleer, M. (1998). Testing for unit roots and non-linear transformations. *Journal of Times Series Analysi*s, *19*, 147--164.
43. Gardner. E. S. Jr. (1985), Exponential smoothing: The state of the art, *Journal of Forecasting*, 4, 1-28.
44. Gardner E. S. Jr. (2006), Exponential smoothing: The state of the art—part II; *International Journal of Forecasting* 22; 636-666.
45. Goldberg, D. E. (1987). Simple genetic algorithms and the minimal deceptive problem, in: L. Davis (Ed.), Genetic Algorithms and Simulated Annealing, Hyperion Books, New York,
46. NY, USA, Morgan Kaufmann, San Francisco, CA, USA, 74–88.
47. Goldberg, D. E. (1989). *Genetic Algorithms, in Search, Optimization & Machine Learning*, Addison-Wesley, Boston, MA, USA
48. Goodwin, P., (2005) How to integrate Management judgment with statistical forecasts. *Foresight*, 1, 8-12.
49. Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, 12, 37—53.
50. Gunst, R.F., and MASON, R.L., (1977) Advantages of examining multi-collinearity in regression Analysis. *Biometrics* 33, 249-260.
51. Gur, Ali. O.., Sayin, S., van Woensel, T., and Fransco, J. (2009). SKU Demand Forecasting in the Presence of Promotions, *Expert Systems with Applications*, 36 (10), 12340-12348.
52. Gur, Ali. O. (2009). Driver-Moderator Method for SKU Sales Forecasting, *in Proceedings of ICDM 2009*.
53. Handzic, M. (1997). Decision Performance as a Function of Information Availability:

An Examination of Executive Information Systems. *Proceedings of the 2nd New South Wales Symposium on Information Technology and Information Systems*, Sydney:UNSW.

54. Hardesty, D. M., & Bearden, W. O. (2003). Consumer Evaluation of Different promotion Types and Price Presentations: The Moderating Role of Promotional Benefit Level, *Journal of retailing*, 79 (1), 17-25.
55. Hardle, W., and Linton, O. (1994). Applied nonparametric Methods, *Handbook of Econometrics*, 4, 2295-2339.
56. Hati, S., Sengupta, S. (2005). Robust camera parameter estimation using genetic algorithm, *Pattern Recognition Letters*, 22 (3/4), 289-298.
57. Heerde, Harald J. Van, Leeflang, Peter S. H., & Wittink, D. R. (2000). The Estimation of Pre- and Post-promotion Dips with Store-Level Scanner Data. *Journal of Marketing Research*, 37 (August), 383-395.
58. Heerde, Harald J. Van, Leeflang, Peter S. H., & Wittink, D. R. (2002a). How promotions work: Scan*Pro-based evolutionary model building. *Schmalenbach Business Review*, 54, 198–220.
59. Heerde, Harald J. Van, Leeflang, Peter S. H., & Wittink, D. R. (2002b). Flexible decomposition of price promotion effects using store-level scanner data.
60. Hendry, D. F. (2000). Econometric modelling. Lecture notes for the PhD course in econometric modelling and economic forecasting, Department of Economics, University of Oslo.
61. Hinterding, R., Michalewicz, Z., and Eiben, A. E. (1997). Adaptation in Evolutionary Computation: A Survey. *IEEE International Conference on Evolutionary Computation. Piscataway*, NJ: IEEE Press. 65-69.
62. Holland, J. H. (1975). Adaptation in natural and artificial systems, University of Michigan Press, Ann Arbor, MI, USA. (extended new edition, MIT Press, Cambridge, MA, USA, 1992).
63. Hyndman, R. J., and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
64. JAMES, W., and STEIN. C. (1961) Estimation with Quadratic Loss. *In proceedings of the fourth Berkeley symposium mathematical statistics and probability*, vol. 1. Berkeley: University of California press, 361-379.
65. Jedidi, K., Mela, C. F., and Gupta, S. (1999). Managing Advertising and Promotion for Long-Run profitability, Marketing Science, 18 (1), 1-22.
66. Klassen, R. D., & Flores, B. E. (2001). Forecasting practices of Canadian firms: Survey results and comparisons. *International Journal of Production Economics*, 70, 163-174.
67. Kleinmutz, B. (1990).Why we still use our heads instead of formulas: towards an integrative approach. *Psychologieal Bulletin*, 107, 296-310.\
68. Kopalle, P. K., Mela, C. F., and Marsh, L. (1999) The dynamic Effect of Discounting on Sales: Empirical Analysis and Normative Pricing Implications. *Marketing Science*, 18(3), 317-332.
69. Kubalik, J., Lazansky, J. (2000). A New Genetic Operator with Enhanced Capabilities.

In: *Quo Vadis Computational Intelligence?*. Sincak, physica-Verlag, Germany, ISBN 3-7908-1324-9, 305-310.

70. Kumar, V., and Peereira, A. (1997) Assessing the competitive impact of type, timing, frequency, and magnitude of retail promotions. *Journal of Business Research* 40, 1-13.
71. Lawrence, M., O'Connor, M. J., Edmundson, R. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operations Research*, 122 (1), 151-160.
72. Lawrence, M., Goodwin, P., O'connor, M., Onkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years, *International Journal of Forecasting*, 22, 493-518.
73. Lee, W. Y., Goodwin, P., and Fildes, R. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, 23, 377-390.
74. Leeflang, P. S. H., and Wittink, D. R. (2000) Building models for marketing decisions: Past, present and future. *International Journal of Research in Marketing*, 17, 105-126.
75. Leung, H. Y., Ching, W. K., and Leung I. K. C. (2008). A Stochastic Optimization Model for Consecutive Promotion, *Quality Technology & Quantitative Management*, 5 (4), 403-414.
76. Libelli, S. M., and Alba, P. (2000). Adaptive Mutation in Genetic Algorithms. *Soft Computing*, 4 (2), 76-80.
77. Lilliefors, H. W. (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *64*, 399-402.
78. Lim, J. S., & O'Connor, M. (1996). Judgmental forecasting with time series and causal information. *International Journal of Forecasting*, 12, 139-153.
79. Lin, J. L., and Liu, T. S. (2002). Modeling lunar calendar holiday effects in Taiwan. *Taiwan Forecasting and Economic Policy Journal*, 33, 1-37.
80. Liu, L. M. (1980). Analysis of time series with calendar effects. *Management Science*, 26 (1), 106-112.
81. Liu, S. H., Mernik, M., Bryant, B. R. (2009). To explore or to exploit: An entropy-driven approach for evolutionary algorithms. *International Journal of Knowledge-based and Intelligent Systems*, 13, 1-14.
82. Liu, Z., Zhou, J., & Lai, S. (2003). New adaptive genetic algorithm based on ranking. *Proceedings of the Second International Conference on Machine Learning and Cybernetics*.
83. Mace, Sandrine., & Neslin, S. A. (2004). The Determinants of Pre- and Postpromotion Dips in Sales of Frequently Purchased Goods. *International Journal of Research in Marketing*, 22(4), 441-457.
84. Makridakis, S., & Maclachlan, D. (1984). Judgmental biases in applied decision making situations. Working Paper, INSEAD, Fountainbleau, France.
85. Mandel, Igor. (2007) Statistical modeling and business expertise, or where is the truth?. Working paper, Advanced marketing Models, 37 Overlook Terrace, Suite 3F New York, NY 10033, USA. igor.mandel@AMModelsInc.com

86. Martinez-Ruiz M.P., Molla-Descals A., Gomez-Borja M.A., Rojo-Alvarez J.L., (2006). Using daily store-level data to understand price promotion effects in a semiparametric regression model, *Journal of Retailing and Consumer Services*, 13, 193-204.
87. Mathews, B. P., & Diamantopoulos, A. (1986). Managerial intervention in forecasting: An empirical investigation of forecast manipulation. *International Journal of Research in Marketing*, 3, 3-10.
88. Meloun, M., and Militky, J. (2001) Detection of single influential points in OLS regression model building. *Analytica Chimica Acta* 439, 169-191.
89. Mentzer, J. T., Bienstock, C. C. (1998). Sales Forecasting Management. Sage, Thousand Oaks, CA.
90. Miller, B. L., and Shaw, M. J. (1996). Genetic algorithms with dynamic niche sharing for multimodal function optimization," in *Proc. 1996 IEEE Int. Conf. Evolutionary Computation*, 1996. Piscataway, NJ: IEEE Press, 786–791.
91. Montero, G., Rodriguez, E., Montenegro, R., Escobar, J. M., Gonzalez-Yuste, J. M. (2005). Genetic algorithms for an improved parameter estimation with local refinement of tetrahedral meshes in a wind model, *Advanced Engineering Software*, 36, 3-10.
92. Narayana, C. L., and Raju, P. S. (1985). Gifts versus sweepstakes: Consumer choices and profiles, *Journal of Advertising*, 14 (1), 50.
93. Nikolopoulos, K., Fildes, R., Goodwin, P., & Lawrence, M. (2005). On the accuracy of judgemental interventionson forecasting support systems. Working paper 005/022. Lancaster University Management School.
94. Nikolopoulos, K., Goodwin, P., Patelis, A., & Assimakopoulos, V. (2006). Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches. *European Journal of Operations Research*, 180, 354--368.
95. O'Connor M., Remus W., and Lim K. (2005). Improving judgmental forecasts with judgmental bootstrapping and task feedback support. *Journal of Behavioral decision making*,18 (4), 247-260.
96. Pepper, J. W. (2010). Efficient Use of Variation in Evolutionary Optimization. *Applied Computational Intelligence and Soft Computing*, 2010, 1-7.
97. Pham, D. T., & Karaboga, D. (1997). Genetic algorithms with variable mutation rates: Application to fuzzy logic controller design. Proceedings of the I MECH E Part I , *Journal of Systems & Control Engineering*, 211(2), 157–167.
98. Raghbir, P. (2004). Free Gift with Purchase: Promoting or Discounting the Brand? *Journal of Consumer Psychology*, 14 (1&2), 181-185.
99. Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied regression Analysis—A Research Tool.* Springer Verlag, New York INC. New York, NY, USA.
100. Remus, W., O'Connor, M., & Griggs, K. (1995). Does reliable information improve the accuracy of judgmental forecasts? *International Journal of Forecasting*, 11, 285-293.
101. Sadownik, R., Barbosa, E. P. (1999). Short-term forecasting of industrial electricity consumption in Brazil, *Journal of Forecasting*, 18 (3), 215-224.
102. Salo, A. A., & Bunn, D. W. (1995). Decomposition in the assessment of judgmental probability forecasts. *Technological forecasting and Social Change*, 149, 13-25.

103. Sanders, N., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making*, 5, 39-52.
104. Sanders, N. R., & Manrodt, K. B. (1994). Forecasting practices in US corporations: Survey results. *Interfaces* 24, 92-100.
105. Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, 31, 511-522.
106. Savin, N.E., and White, K.J. (1977) The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors', *Ecometrica* 45, 1989-1996.
107. Scapolo, F., & Miles, Ian. (2006). Eliciting experts' knowledge: A comparison of two methods. *Technological Forecasting & Social Change*, 73, 679—704.
108. Schaffer, J. D., Caruana, R. A., Eshelman, L. J., & Das, R. (1989). A study of control parameters affecting online performance of genetic algorithms for function optimization. In Proceedings of the Third International Conference on Genetic Algorithms, 51–60.
109. Schoenauer, M., and Sebag, M. (2002). Using Domain Knowledge in Evolutionary System Identification.InK.G.etal.,editor, Evolutionary Algorithms in Engineering and Computer Science.JohnWiley.
110. Shadownik, R., Barbosa, E.P., (1999). Short-term forecasting of industrial electricity consumption in Brazil. *Journal of Forecasting*, 18 (3), 215-224.
111. Singh, G. and Deb, K. (2006). Comparison of multi-modal optimization algorithms based on evolutionary algorithms. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-06), 1305–1312.
112. Singh, N., and Takhtani, A. (2005). Weekend rush blues, *Times of India Mumbai*, Consumer Edge, 16-17.
113. Skalak, D. B. (1994). Prototype and feature selection by sampling and random mutation hill climbing algorithms. In Proceedings of the Eleventh International Conference on Machine Learning, (pp. 293--301), New Brunswick, NJ. Morgan Kaufmann.
114. Slinker, B.K., and Glantz, S.A. (1985) Multiple regression for physiological data analysis: the problem of multi-collinearity. *American journal of Physiology* 249, 1-12.
115. Smith, G., and Campbell, F. (1980) A critique of some ridge regression methods. *Journal of American Statistics Association.* 75, 369, 74-81.
116. Spears, W. M. (1992). Adapting crossover in a genetic algorithm. Naval Research Laboratory AI Center Report AIC-92-025. Washington, DC 20375 USA.
117. Stewart T. R., and C.M. Lusk., (1994). Seven Components of Judgmental Forecasting Skill: Implications for Research and the Improvement of Forecasts, *Journal of Forecasting*, 13, 579-599.
118. Stine, R.A. (1995) Graphical interpretation of variance inflation factors. *The American Statistician* 45, 53-56.
119. Soukup, R. J., and Findley D. F. (2000). Detection and Modeling of Trading Day Effects, in ICES II: Proceedings of the Second International Conference on Establishment Surveys (2001),743-753, Alexandria: American Statistical Association.
120. Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., & Goodwin, P. (2009). The effects of integreating management judgement into intermittent demand forecasts.

International Journal of Production Economics, 118, 72-81.
121. Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend, *International Journal of Forecasting*, 19, 715-725.
122. Teasley, B. Marketing Mix Modeling. www.teasley.net
123. Tellis, G. J., Zufryden, F. S. Tackling the Retailer Decision Maze: Which Brands to Discount, How Much, When and Why? *Marketing Science*, 14 (3), 271-299.
124. Theil, H. (1971) *Principles of Econometrics*. New York: Wiley.
125. Thonemann, U.W. (2002). Improving Supply-chain Performance by Sharing Advance Demand Information, *European Journal of Operational Research* 142, 81--107.
126. Tversky, A., and Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Scienc*e, 185, 1124-1131.
127. Trusov, M., Bodapati, A. V., and Cooper, L. G. (2006). Retailer promotion planning: Improving forecast accuracy and interpretability. *Journal of Interactive Marketing*, 20 (3-4), 71-81.
128. Wang, C. L., & Wang, L. C. (2010). A GA-based forecasting model focused on contextual factors. In Proceedings of the International Conference on Knowledge Globalization 2010, 21-29.
129. Webby, R., O'Connor, M., and Edmundson, B. (2005). Forecasting support systems for the incorporation of event information: An empirical investigation, *International Journal of Forecasting*, 21 (3), 411-423.
130. White, C. M.,Yen, .G. G. (2004). A hybrid evolutionary algorithm for traveling salesman problem, Congress on Evolutionary Computation,Vol.2,1473-1478.
131. Williams, T. M. (1987). Adaptive Holt-Winters forecasting. *Journal of Operational Research Society*, *38 (6)*, 553-560.
132. Xie, J., and Xing, W. (1998). Incorporating Domain-Specific Knowledge into Evolutionary Algorithms, *Beijing Mathematic*s, Vol.4, No.2, 131-139.
133. Xu, Z., and Gao, Y. (1997). Characteristic analysis and prevention on premature convergence in genetic algorithms. Science of China, Vol. 40, No.2, 113-125.
134. Yu, C.H. (2000). An overview of remedial tools for collinearity in SAS. In proceedings of 2000 Western Users of SAS Software Conference, 196-201.
135. Zhu, K., Thonemann, U.W., (2004). Modeling the Benefits of Sharing Future Demand Information, *Operations Research* 52, 136-147.

Appendix

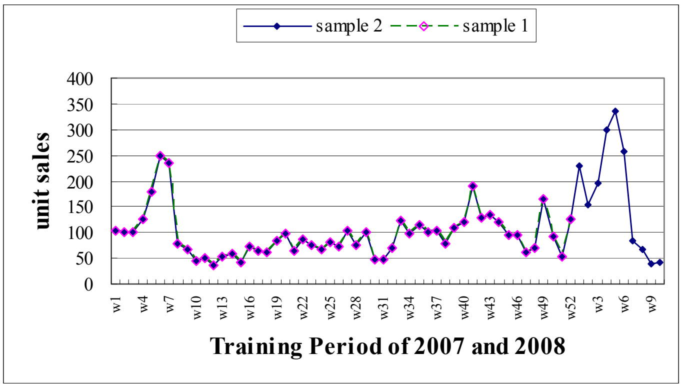

Figure A1. Sales data of item 1 in training period of 2007 and 2008

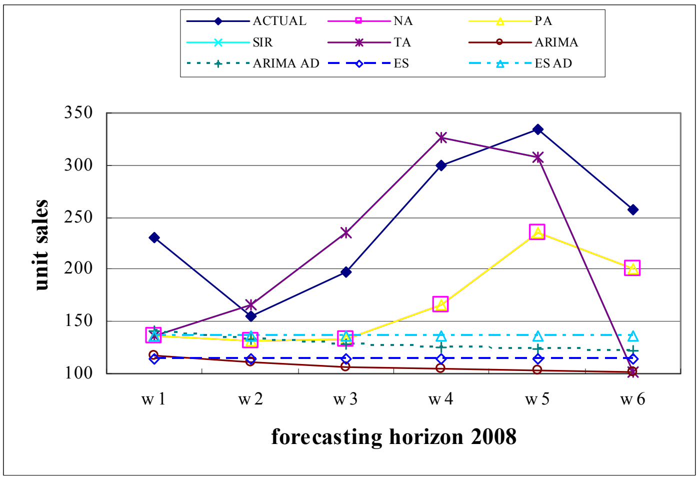

Figure B1. A comparison of various adjustment methods for item 1 in busy period of 2008

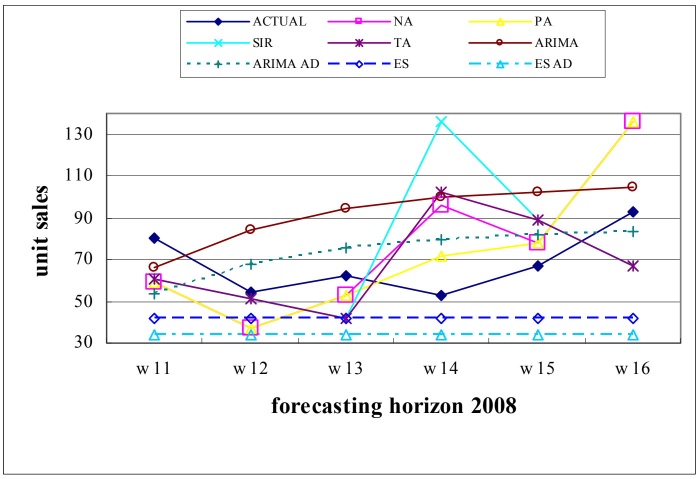

Figure C1. A comparison of various adjustment methods for item 1 in off period of 2008

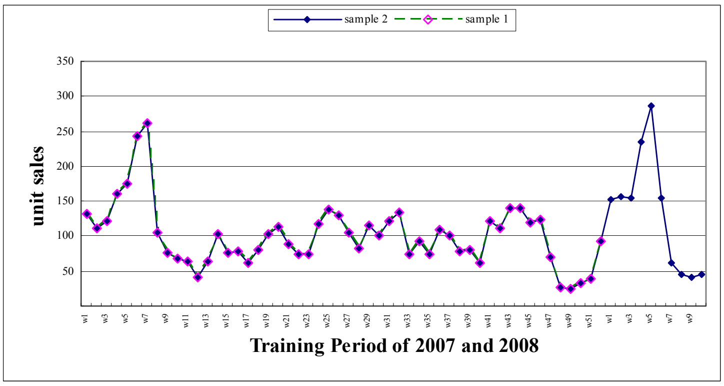

Figure A2. Sales data of item 2 in training period of 2007 and 2008

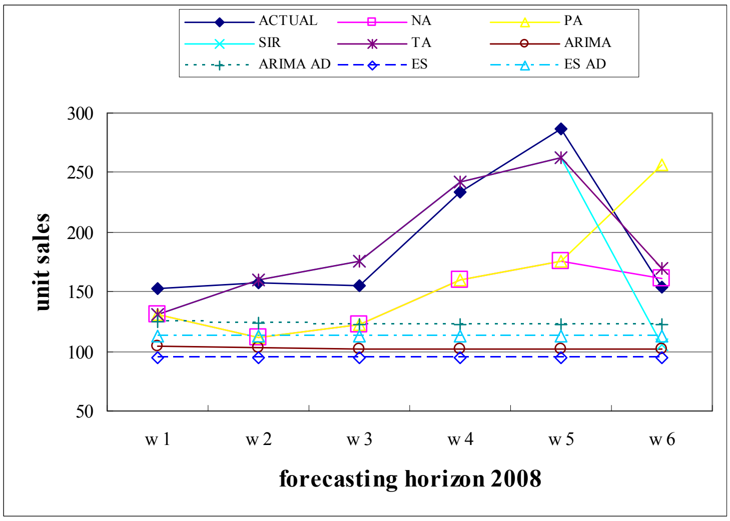

Figure B2. A comparison of various adjustment methods for item 2 in busy period of 2008

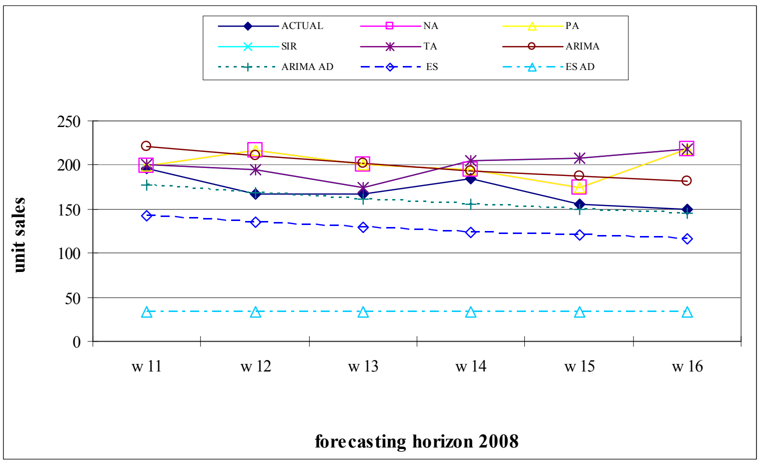

Figure C2. A comparison of various adjustment methods for item 2 in off period of 2008

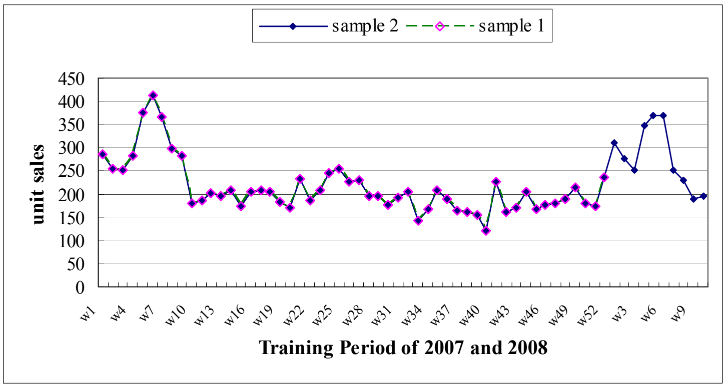

Figure A3. Sales data of item 3 in training period of 2007 and 2008

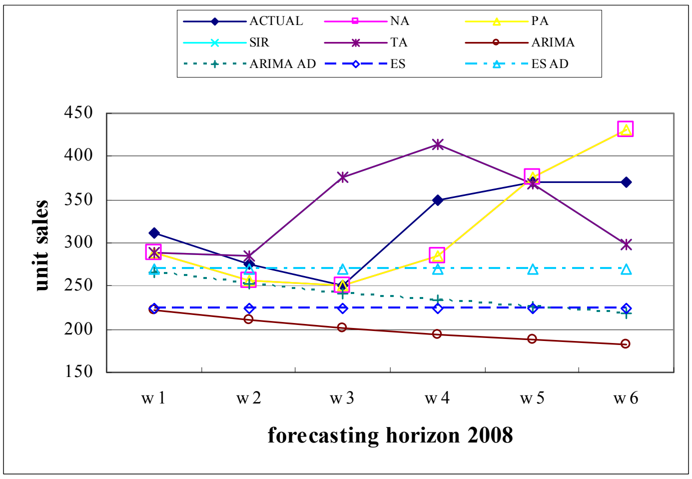

Figure B3. A comparison of various adjustment methods for item 3 in busy period of 2008

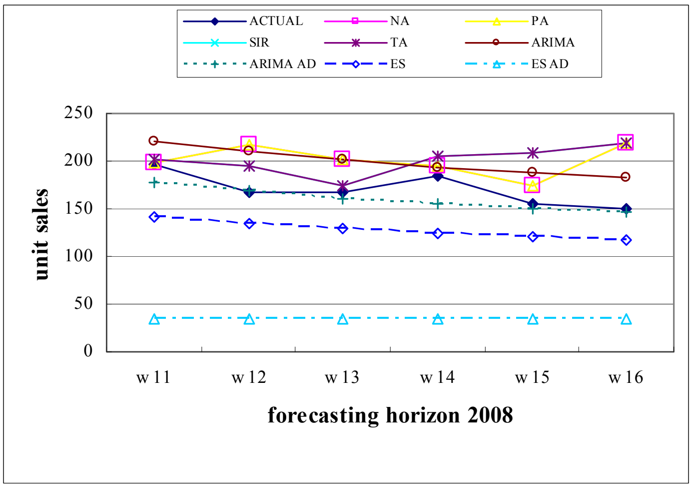

Figure C3. A comparison of various adjustment methods for item 3 in off period of 2008

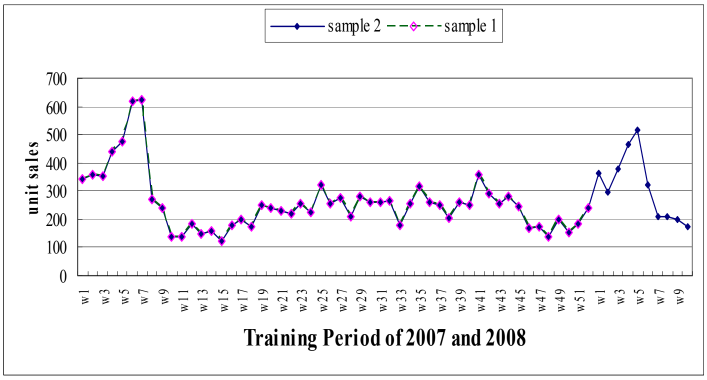

Figure A4. Sales data of item 4 in training period of 2007 and 2008

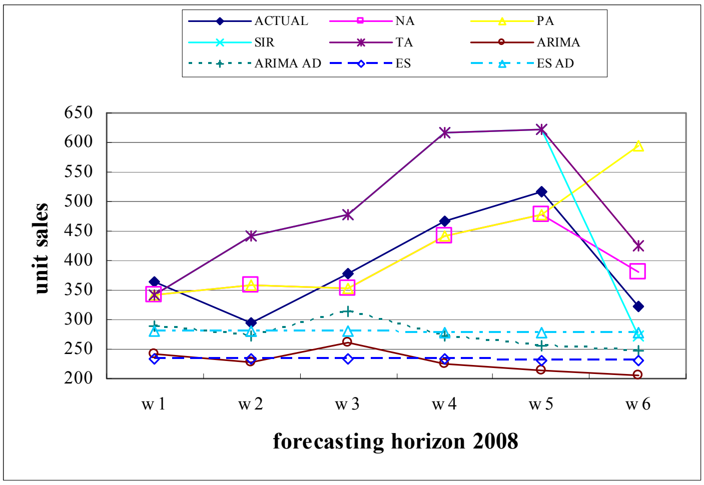

Figure B4. A comparison of various adjustment methods for item 4 in busy period of 2008

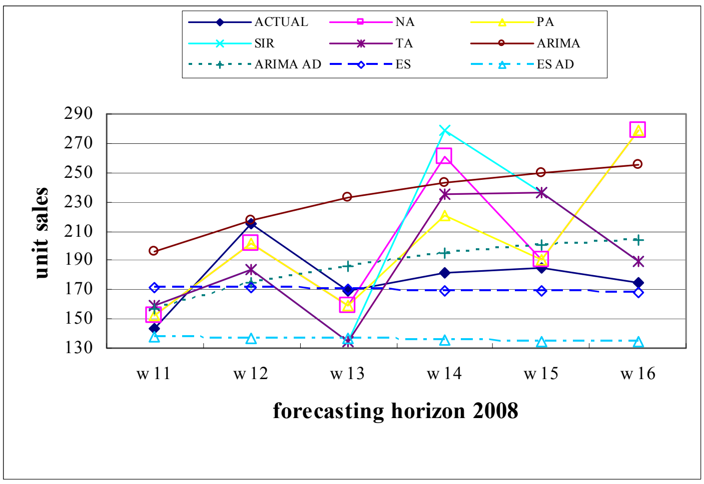

Figure C4. A comparison of various adjustment methods for item 4 in off period of 2008

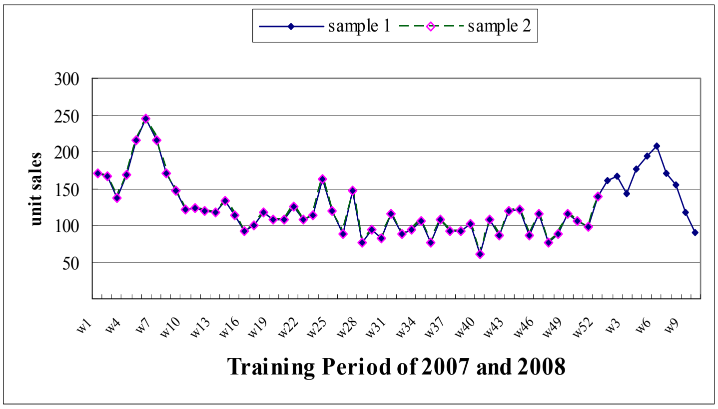

Figure A5. Sales data of item 5 in training period of 2007 and 2008

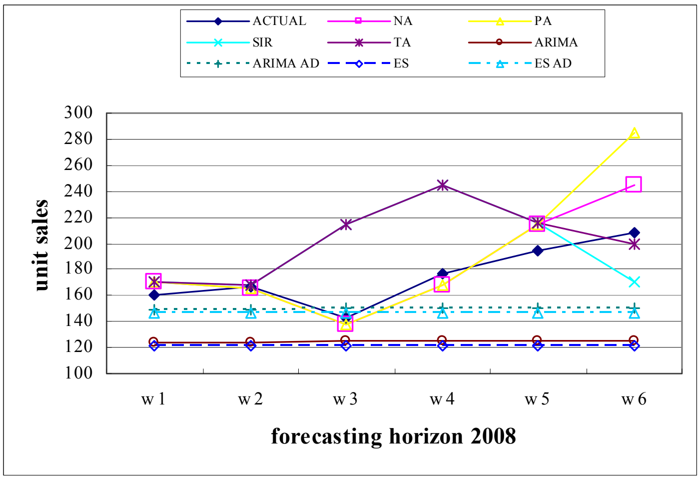

Figure B5. A comparison of various adjustment methods for item 5 in busy period of 2008

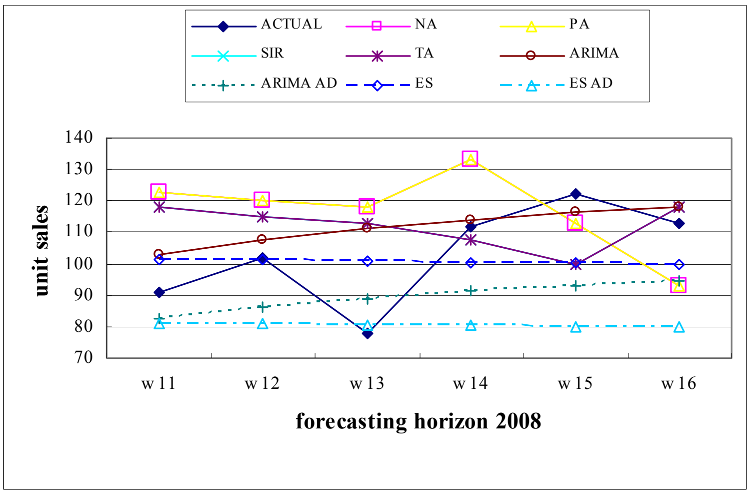

Figure C5. A comparison of various adjustment methods for item 5 in off period of 2008

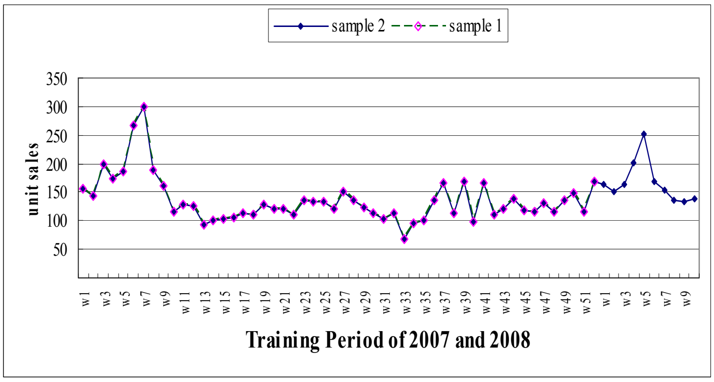

Figure A6. Sales data of item 6 in training period of 2007 and 2008

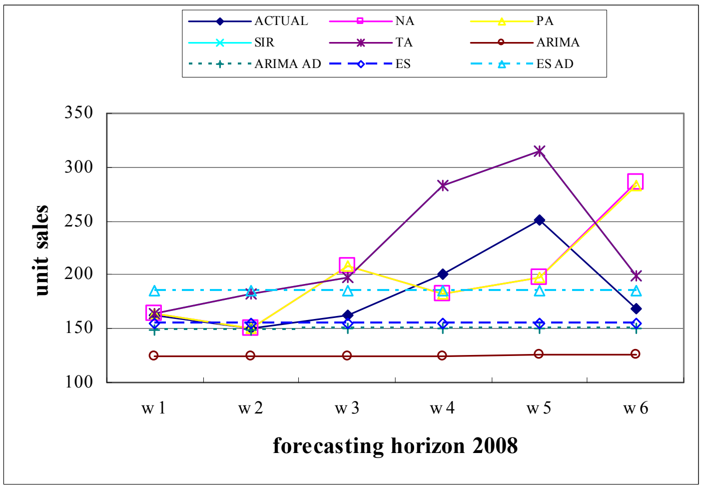

Figure B6. A comparison of various adjustment methods for item 6 in busy period of 2008

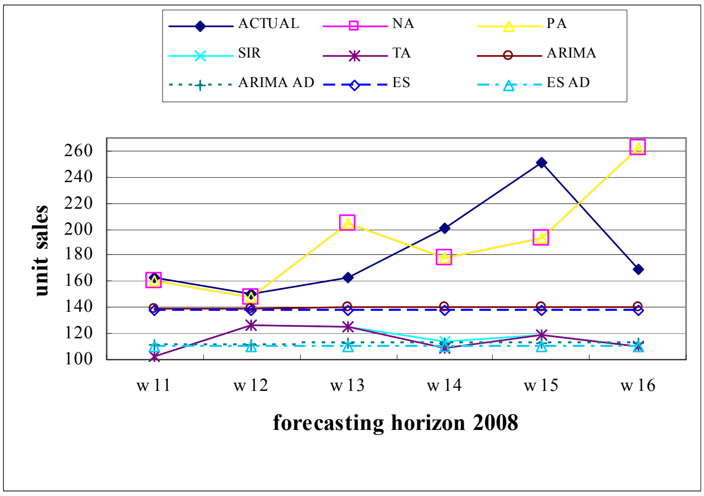

Figure C6. A comparison of various adjustment methods for item 6 in off period of 2008

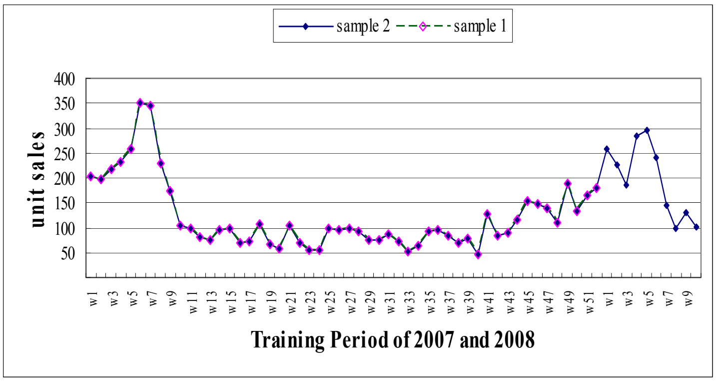

Figure A7. Sales data of item 7 in training period of 2007 and 2008

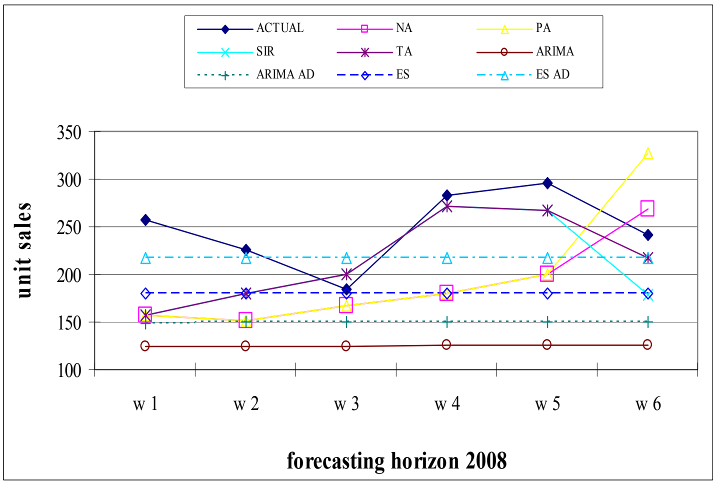

Figure B7. A comparison of various adjustment methods for item 7 in busy period of 2008

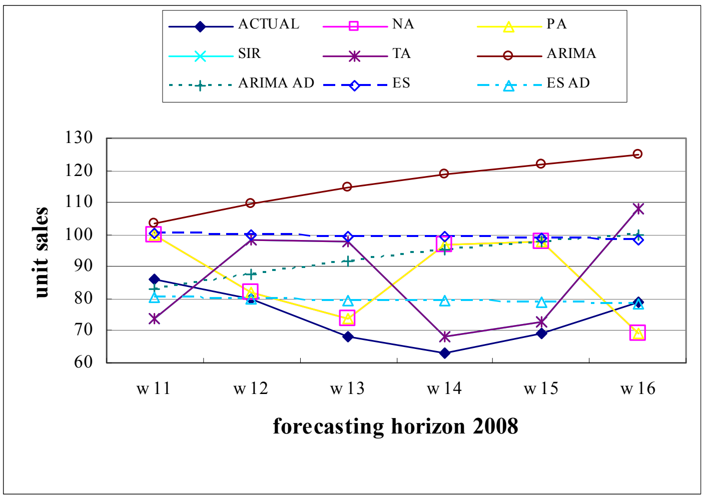

Figure C7. A comparison of various adjustment methods for item 7 in off period of 2008

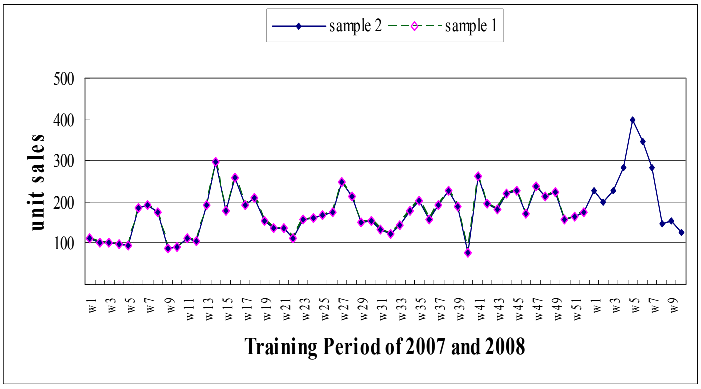

Figure A8. Sales data of item 8 in training period of 2007 and 2008

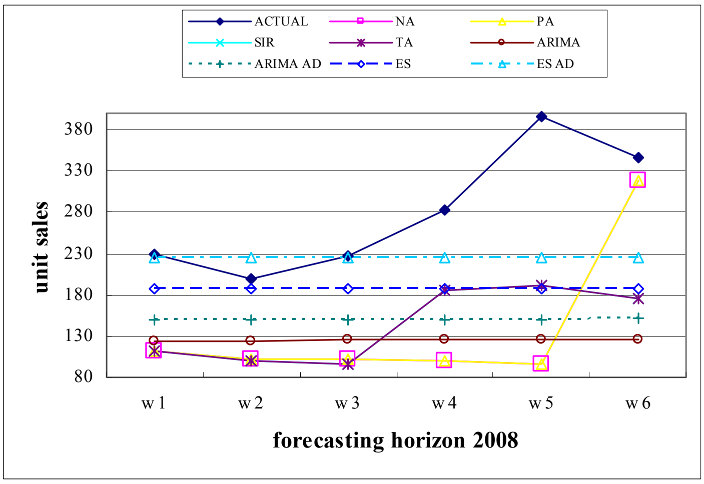

Figure B8. A comparison of various adjustment methods for item 8 in busy period of 2008

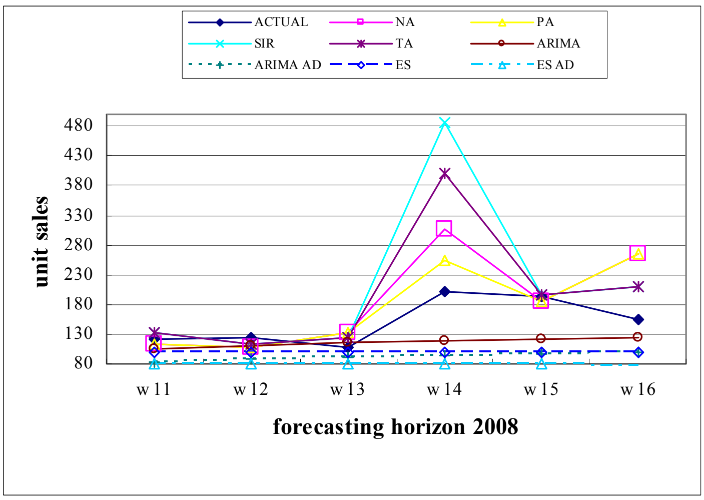

Figure C8. A comparison of various adjustment methods for item 8 in off period of 2008

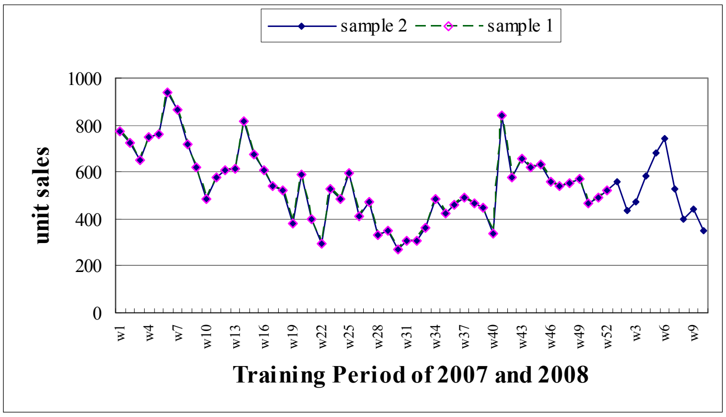

Figure A9. Sales data of item 9 in training period of 2007 and 2008

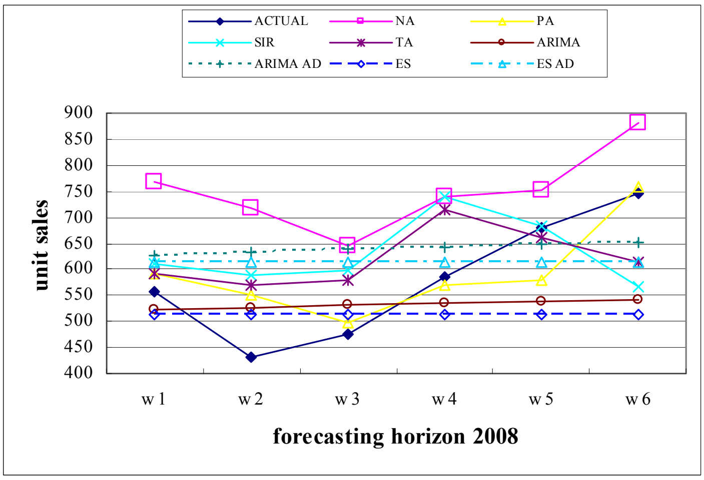

Figure B9. A comparison of various adjustment methods for item 9 in busy period of 2008

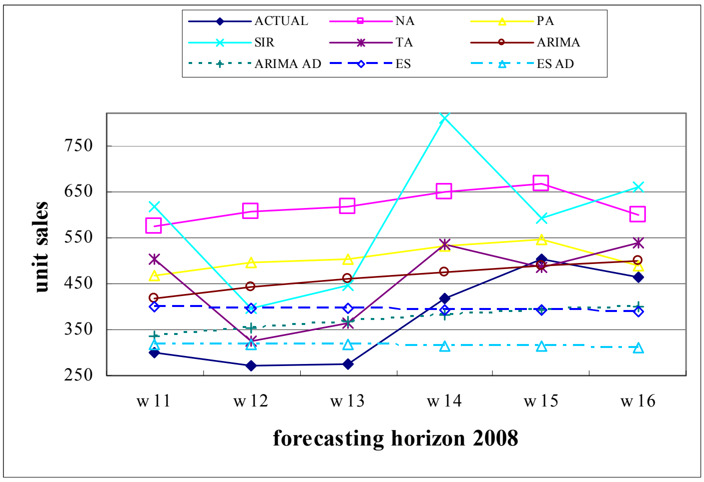

Figure C9. A comparison of various adjustment methods for item 9 in off period of 2008

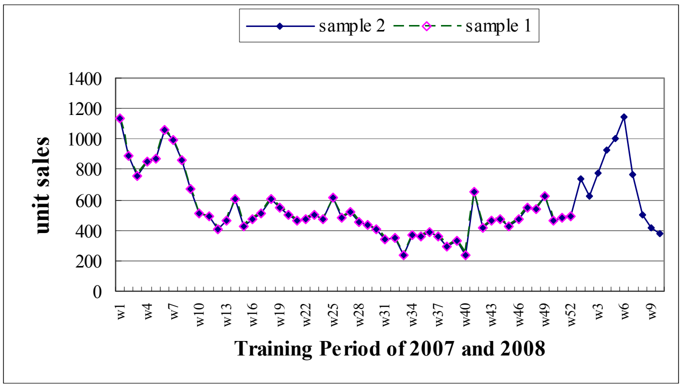

Figure A10. Sales data of item 10 in training period of 2007 and 2008

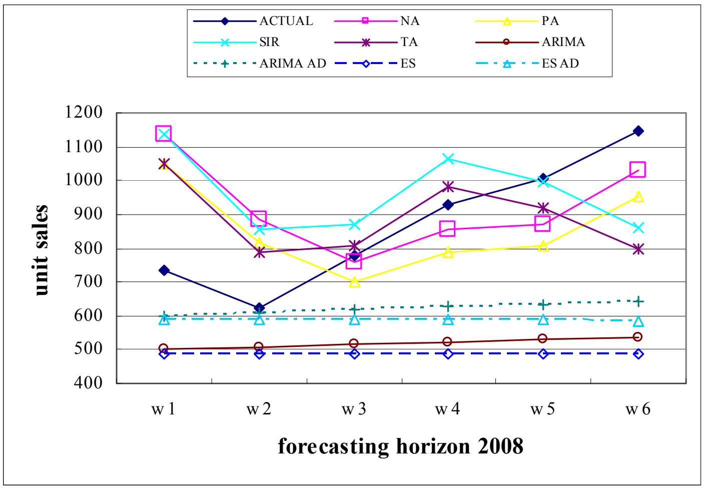

Figure B10. A comparison of various adjustment methods for item 10 in busy period of 2008

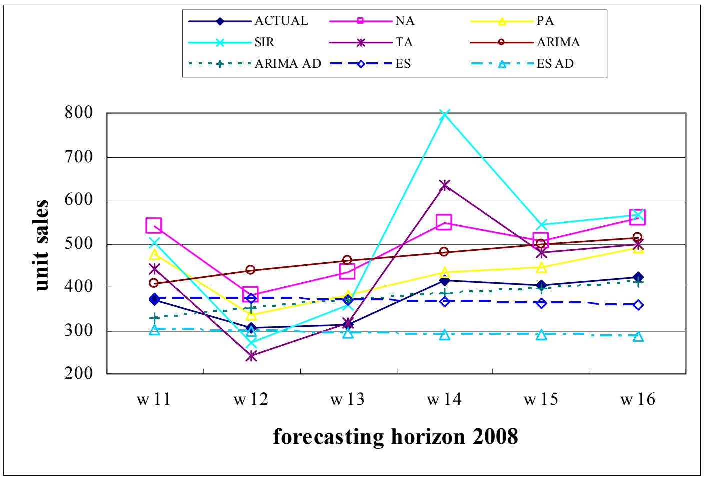

Figure C10. A comparison of various adjustment methods for item 10 in off period of 2008

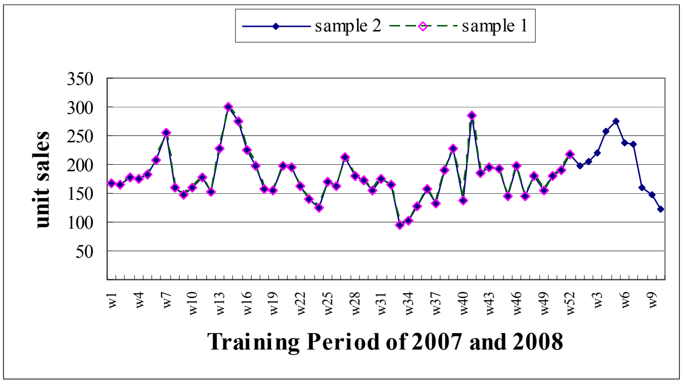

Figure A11. Sales data of item 11 in training period of 2007 and 2008

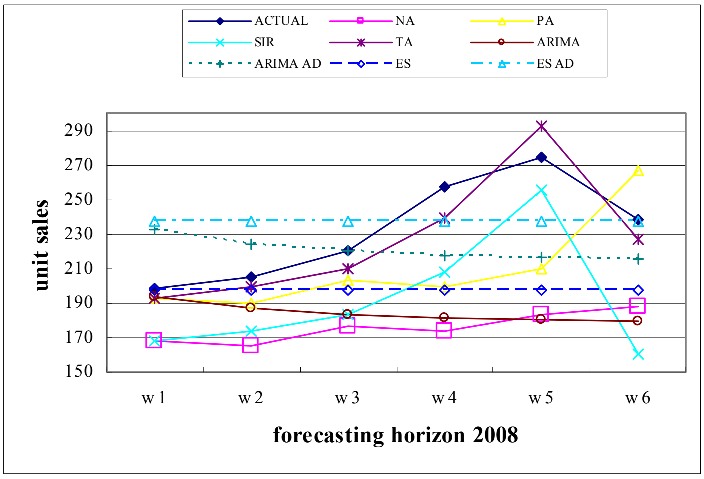

Figure B11. A comparison of various adjustment methods for item 11 in busy period of 2008

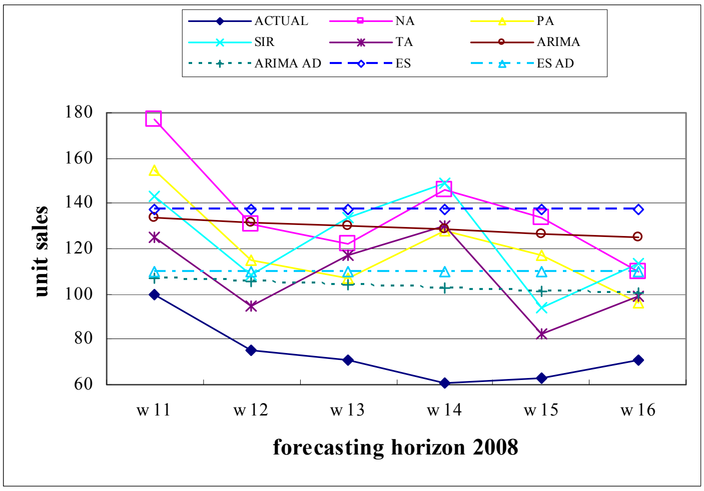

Figure C11. A comparison of various adjustment methods for item 11 in off period of 2008

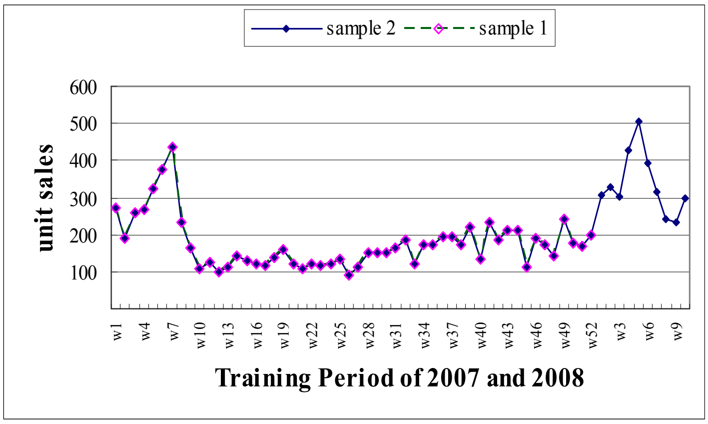

Figure A12. Sales data of item 12 in training period of 2007 and 2008

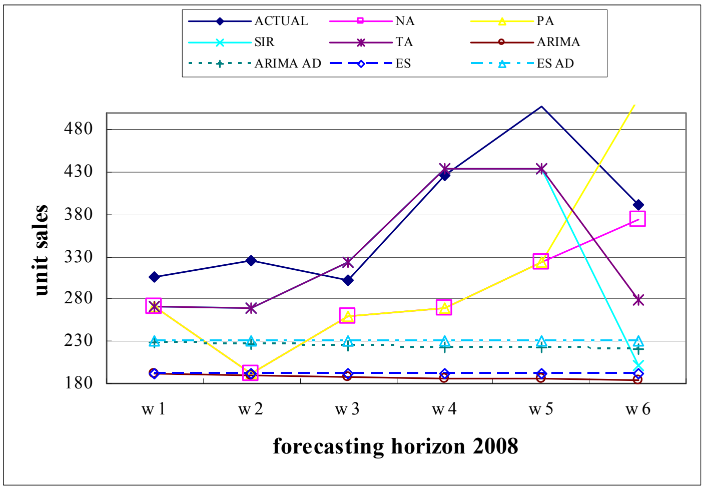

Figure B12. A comparison of various adjustment methods for item 12 in busy period of 2008

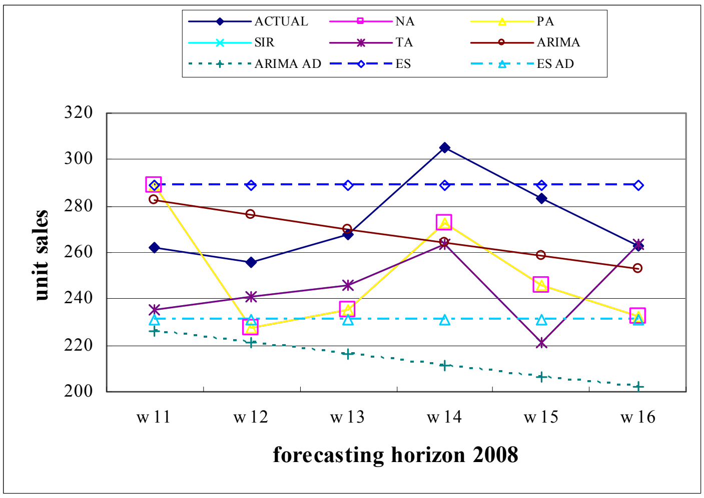

Figure C12. A comparison of various adjustment methods for item 12 in off period of 2008

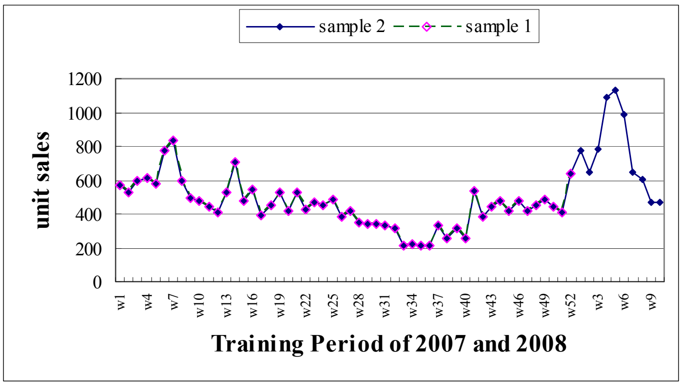

Figure A13. Sales data of item 13 in training period of 2007 and 2008

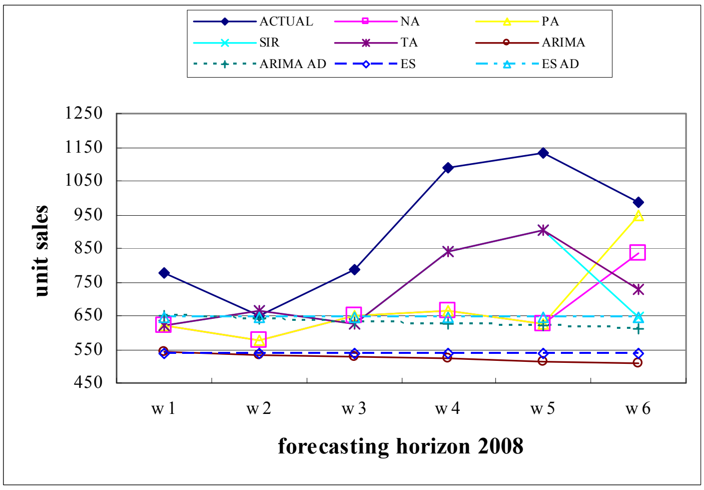

Figure B13. A comparison of various adjustment methods for item 13 in busy period of 2008

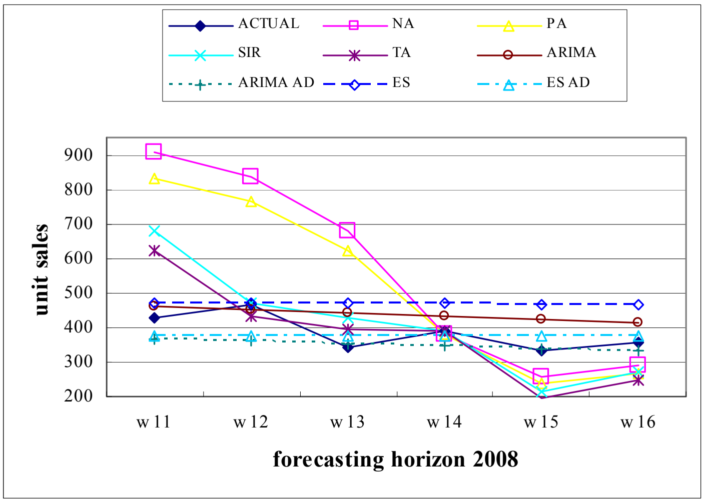

Figure C13. A comparison of various adjustment methods for item 13 in off period of 2008

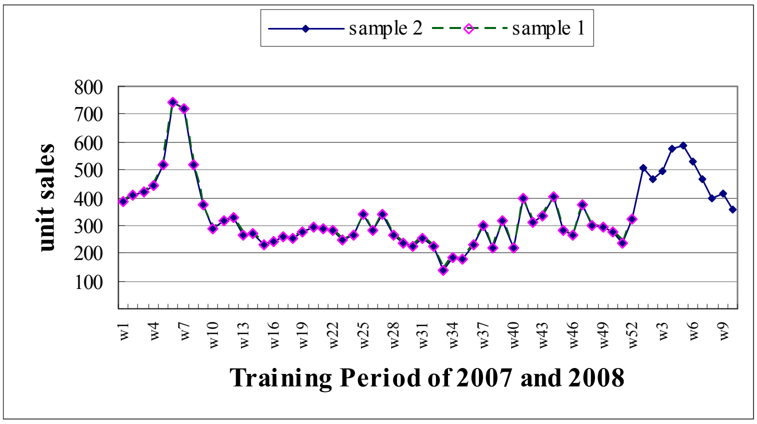

Figure A14. Sales data of item 14 in training period of 2007 and 2008

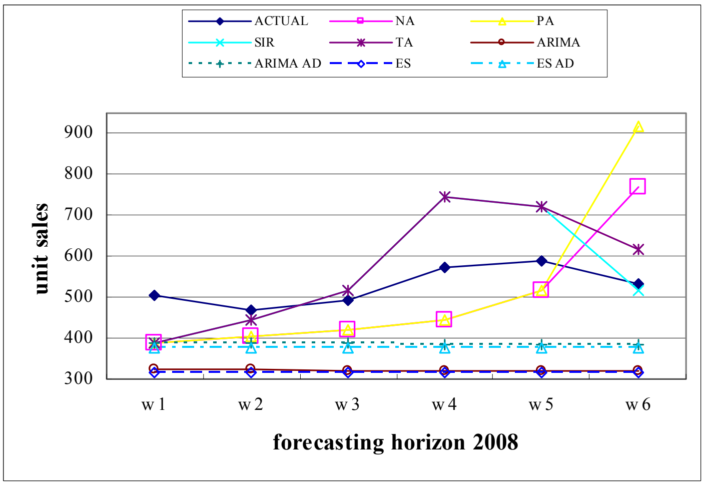

Figure B14. A comparison of various adjustment methods for item 14 in busy period of 2008

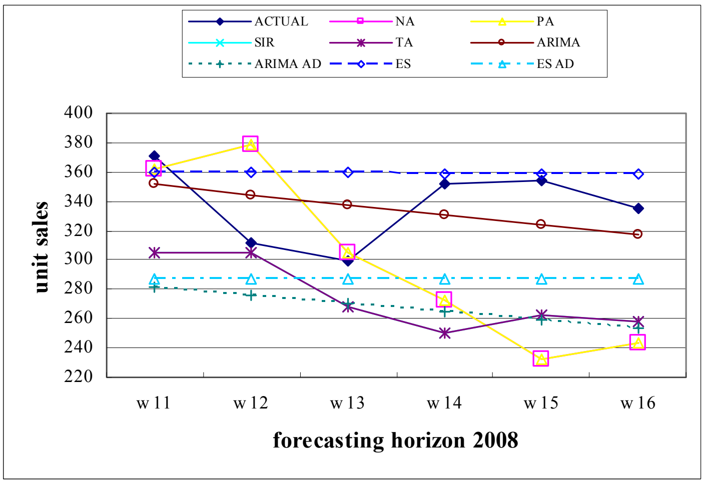

Figure C14. A comparison of various adjustment methods for item 14 in off period of 2008

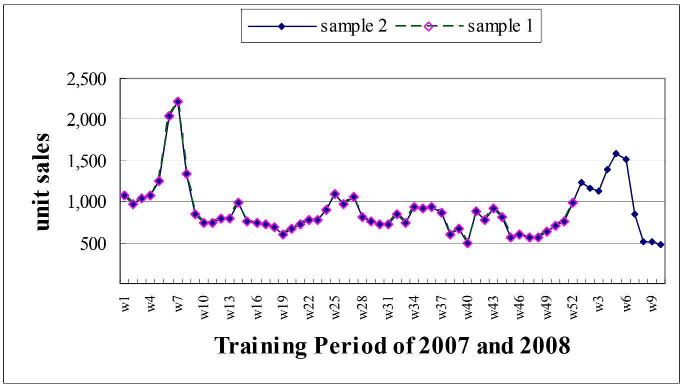

Figure A15. Sales data of item 15 in training period of 2007 and 2008

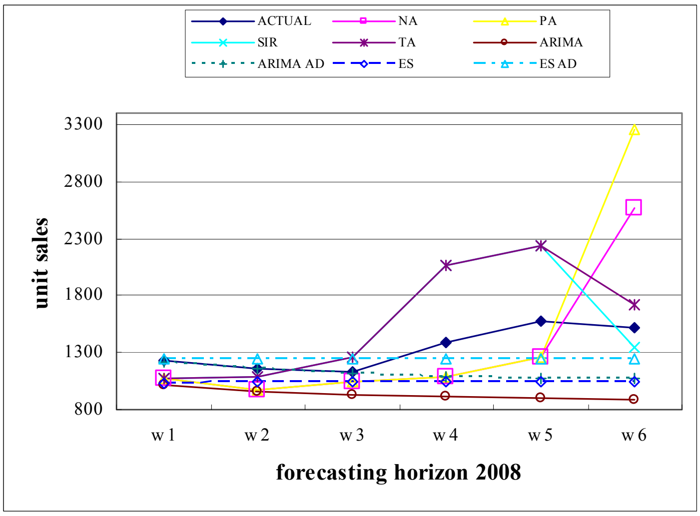

Figure B15. A comparison of various adjustment methods for item 15 in busy period of 2008

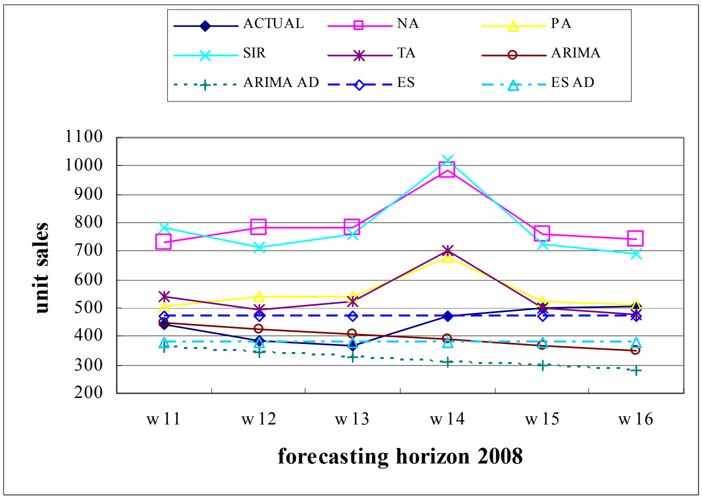

Figure C15. A comparison of various adjustment methods for item 15 in off period of 2008

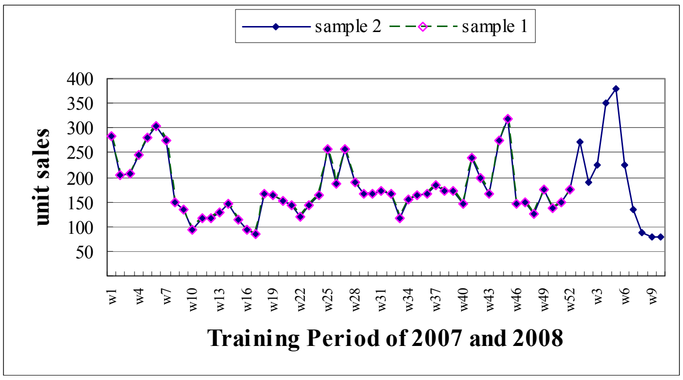

Figure A16. Sales data of item 16 in training period of 2007 and 2008

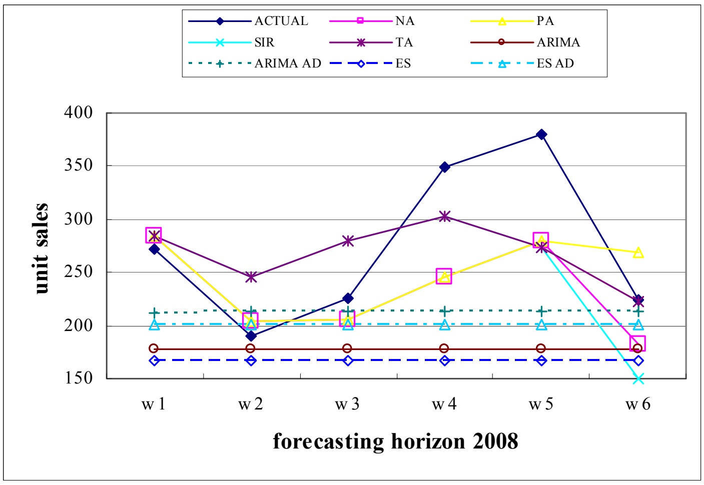

Figure B16. A comparison of various adjustment methods for item 16 in busy period of 2008

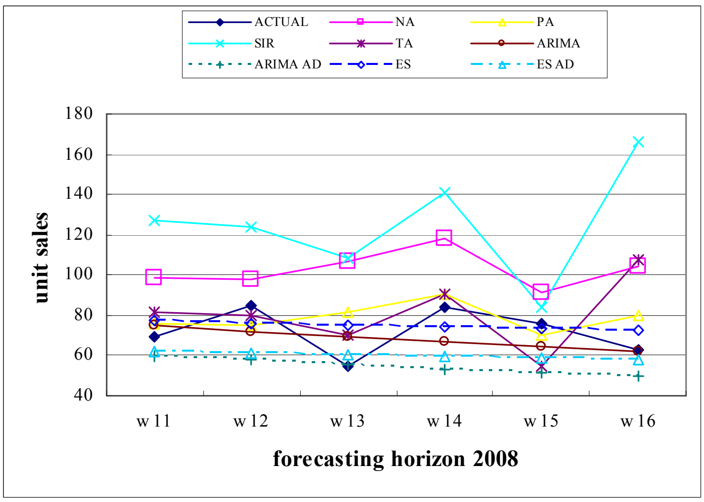

Figure C16. A comparison of various adjustment methods for item 16 in off period of 2008

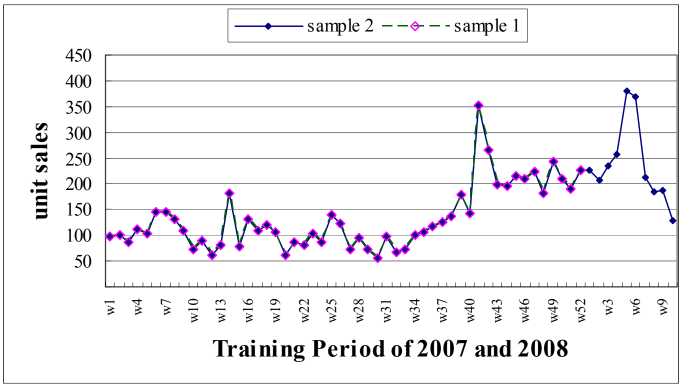

Figure A17. Sales data of item 17 in training period of 2007 and 2008

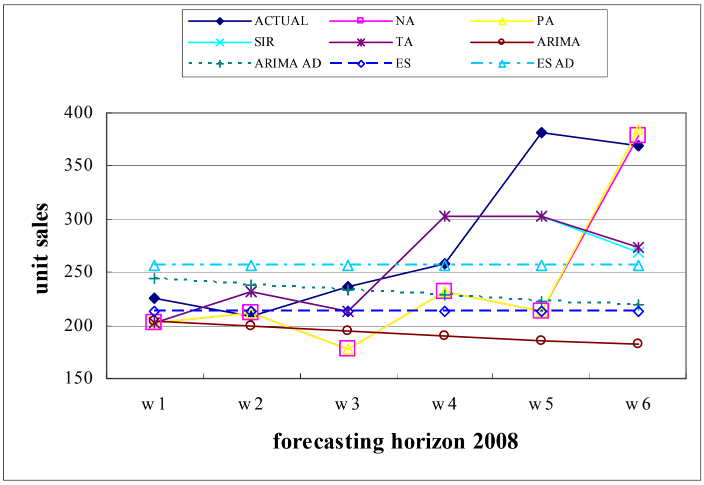

Figure B17. A comparison of various adjustment methods for item 17 in busy period of 2008

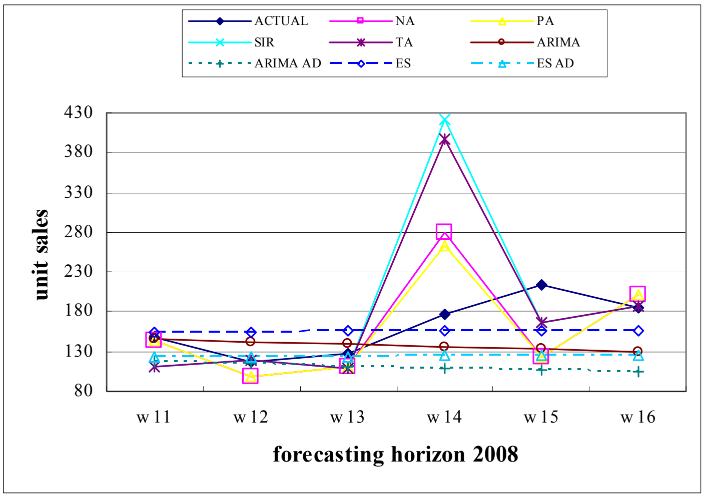

Figure C17. A comparison of various adjustment methods for item 17 in off period of 2008

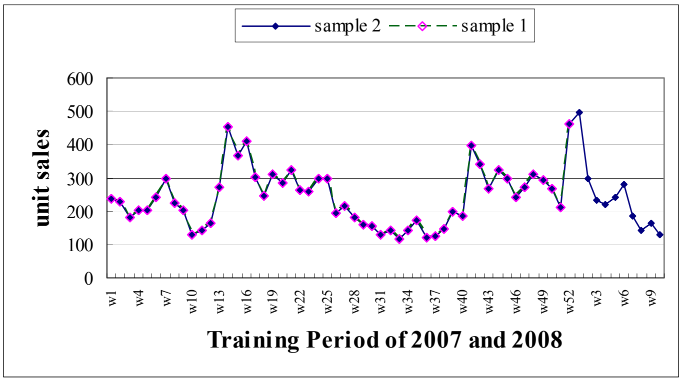

Figure A18. Sales data of item 18 in training period of 2007 and 2008

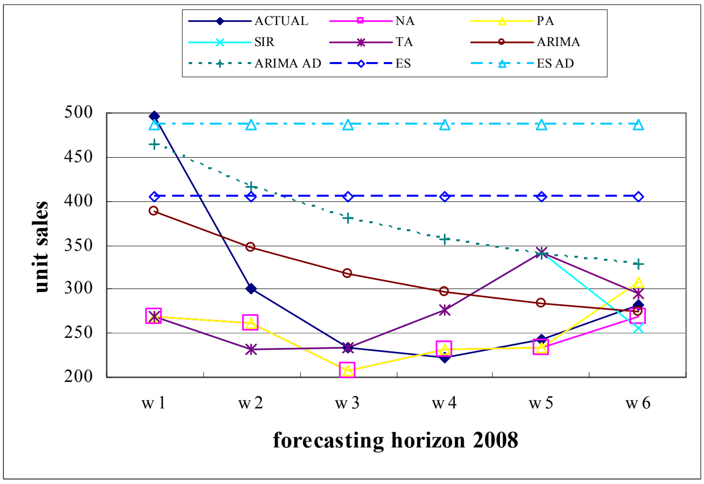

Figure B18. A comparison of various adjustment methods for item 18 in busy period of 2008

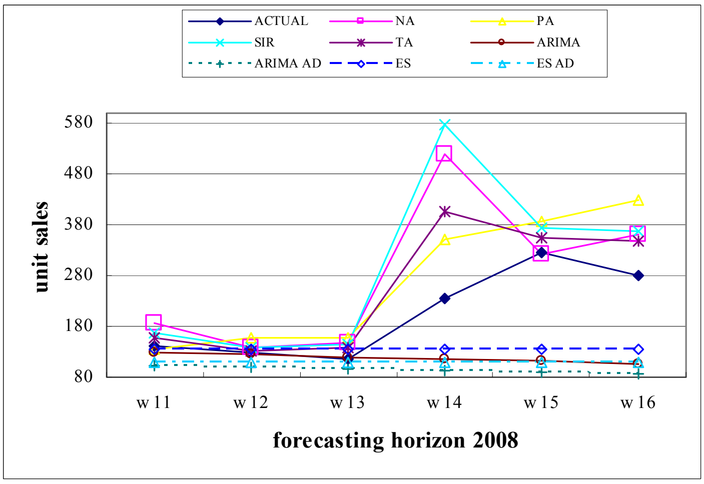

Figure C18. A comparison of various adjustment methods for item 18 in off period of 2008

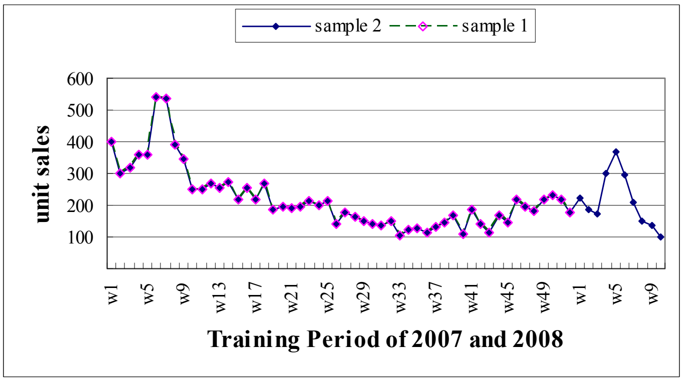

Figure A19. Sales data of item 19 in training period of 2007 and 2008

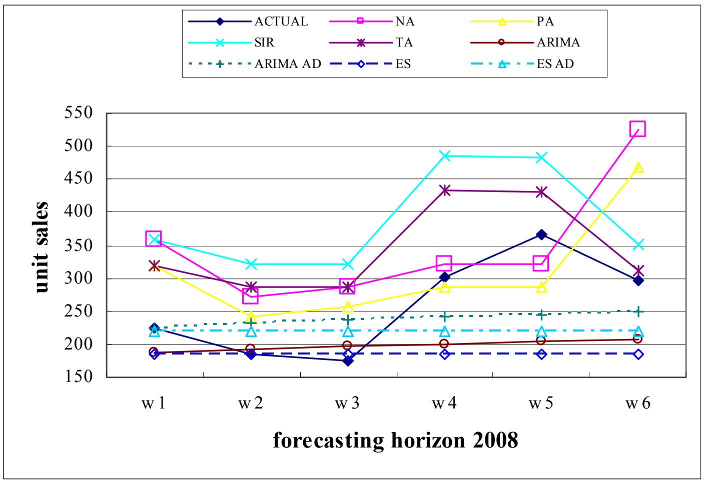

Figure B19. A comparison of various adjustment methods for item 19 in busy period of 2008

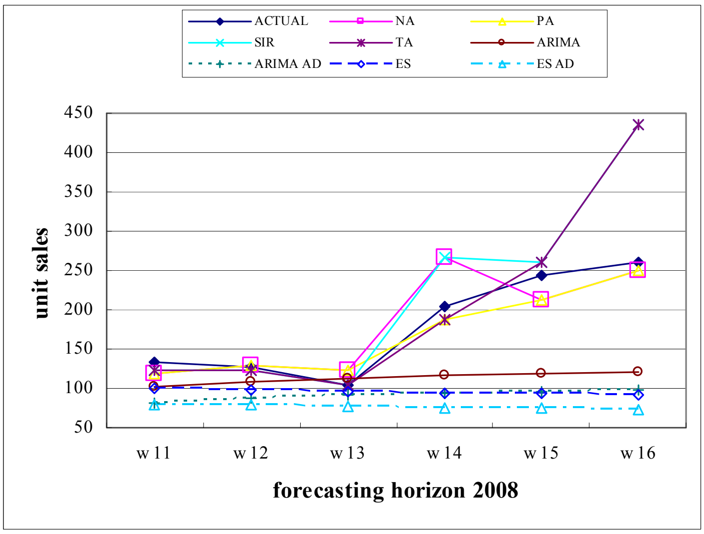

Figure C19. A comparison of various adjustment methods for item 19 in off period of 2008

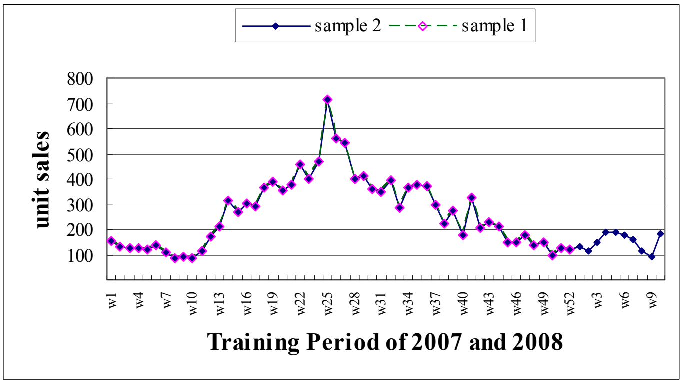

Figure A20. Sales data of item 20 in training period of 2007 and 2008

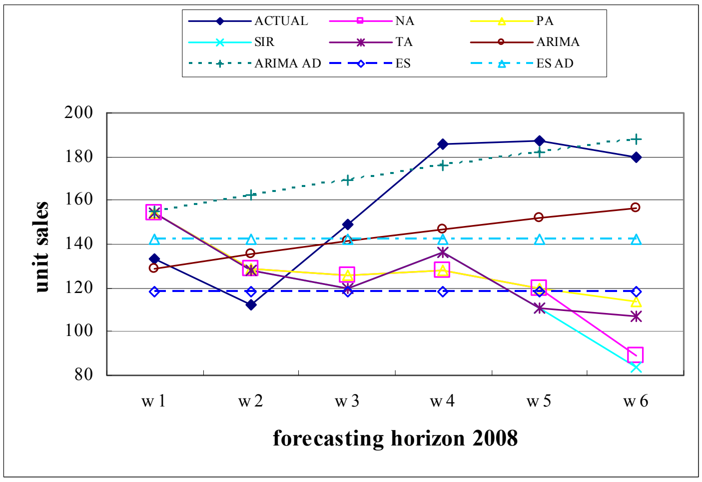

Figure B20. A comparison of various adjustment methods for item 20 in busy period of 2008

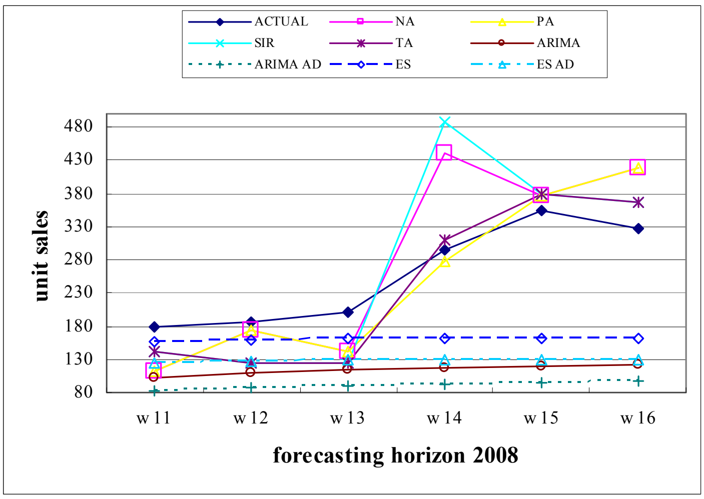

Figure C20. A comparison of various adjustment methods for item 20 in off period of 2008

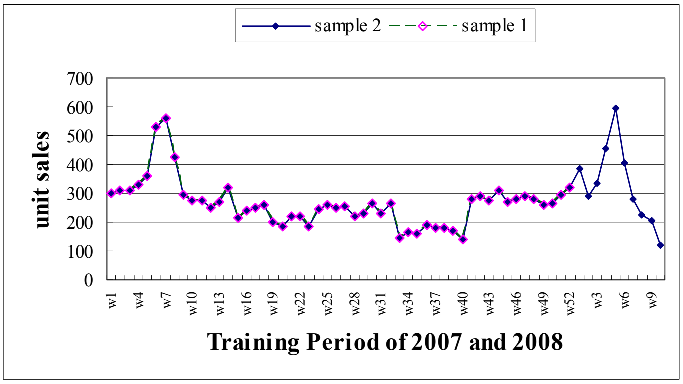

Figure A21. Sales data of item 21 in training period of 2007 and 2008

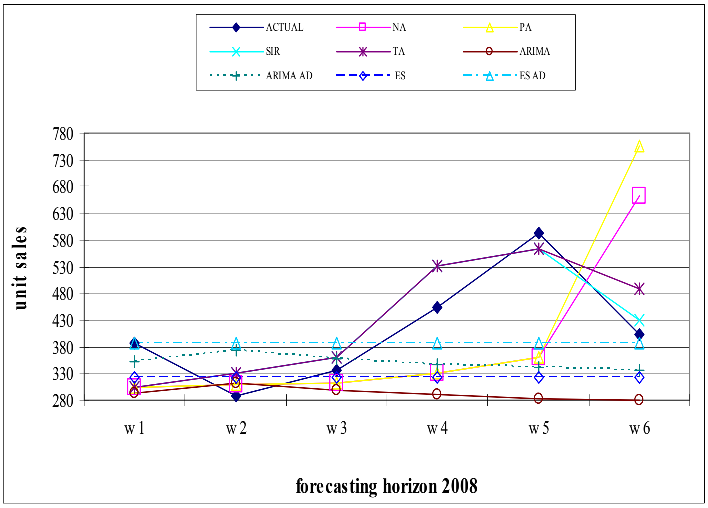

Figure B21. A comparison of various adjustment methods for item 21 in busy period of 2008

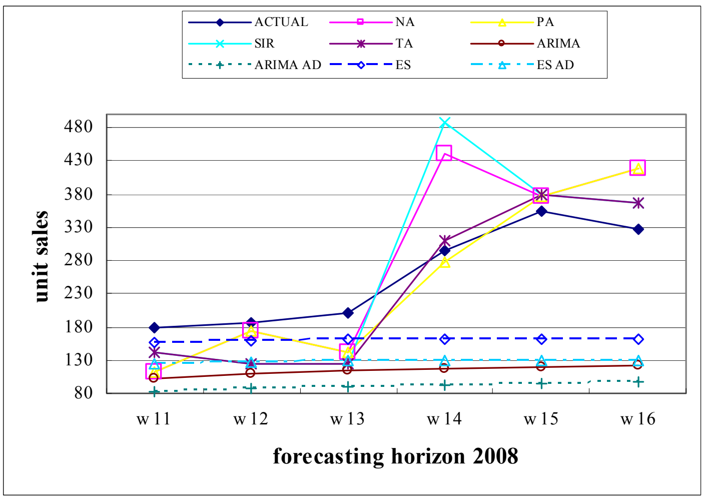

Figure C21. A comparison of various adjustment methods for item 21 in off period of 2008

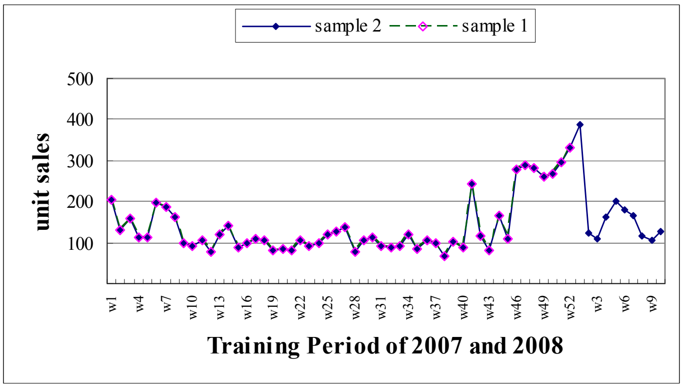

Figure A22. Sales data of item 22 in training period of 2007 and 2008

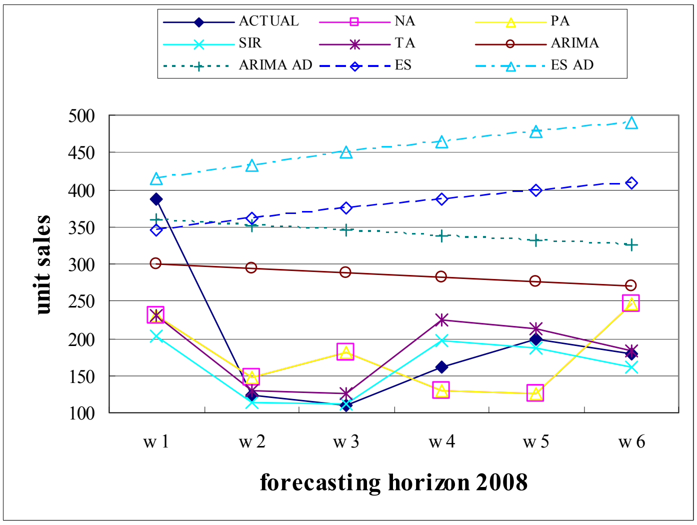

Figure B22. A comparison of various adjustment methods for item 22 in busy period of 2008

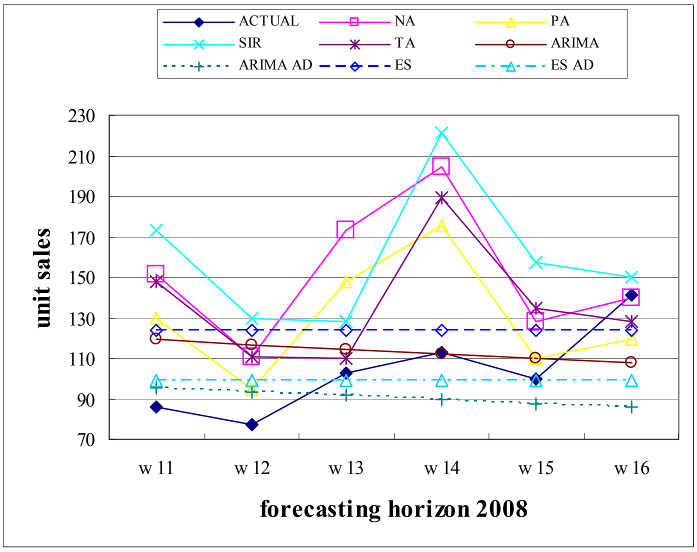

Figure C22. A comparison of various adjustment methods for item 22 in off period of 2008

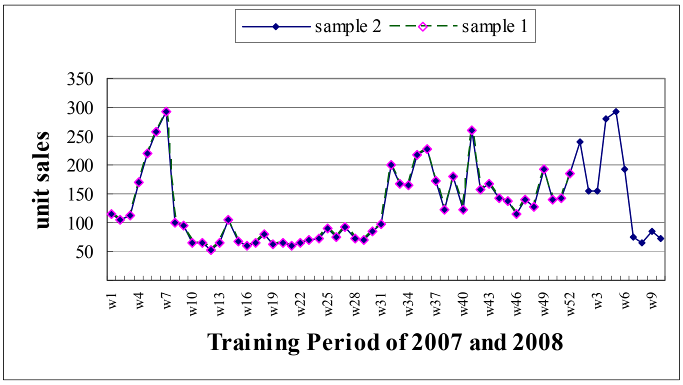

Figure A23. Sales data of item 23 in training period of 2007 and 2008

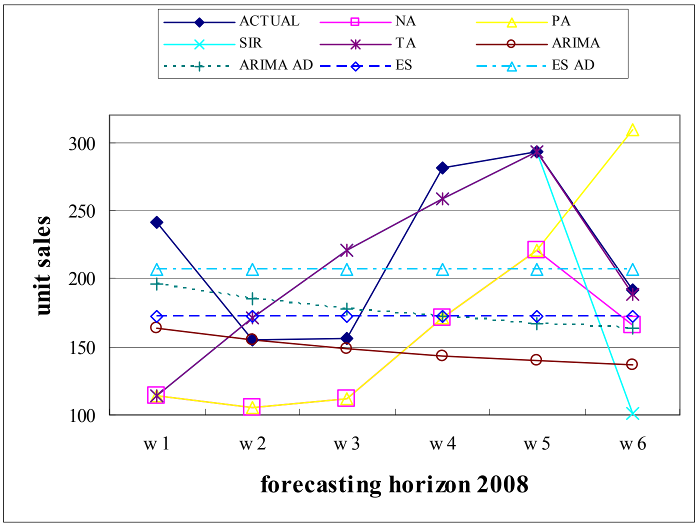

Figure B23. A comparison of various adjustment methods for item 23 in busy period of 2008

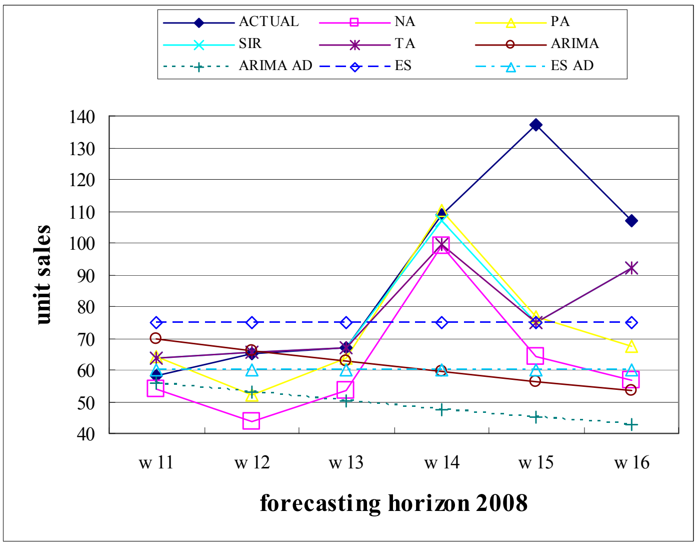

Figure C23. A comparison of various adjustment methods for item 23 in off period of 2008

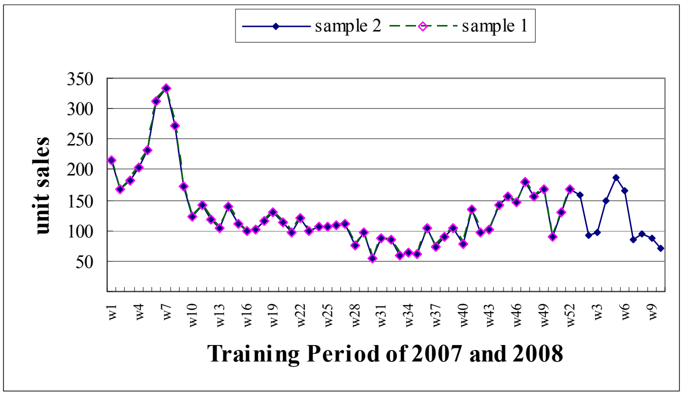

Figure A24. Sales data of item 24 in training period of 2007 and 2008

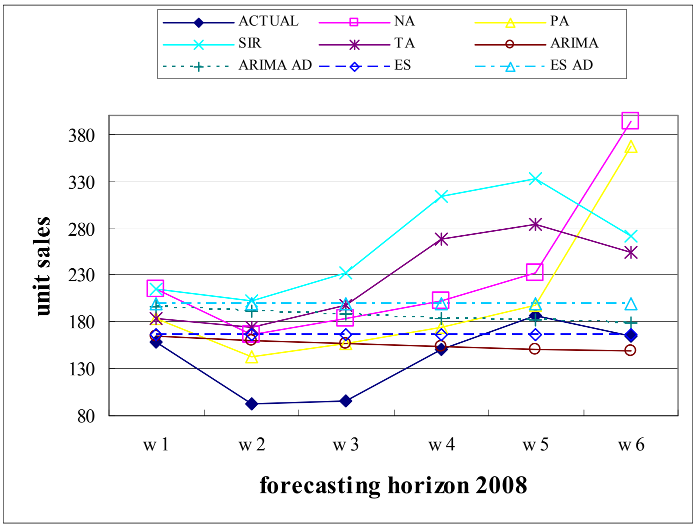

Figure B24. A comparison of various adjustment methods for item 24 in busy period of 2008

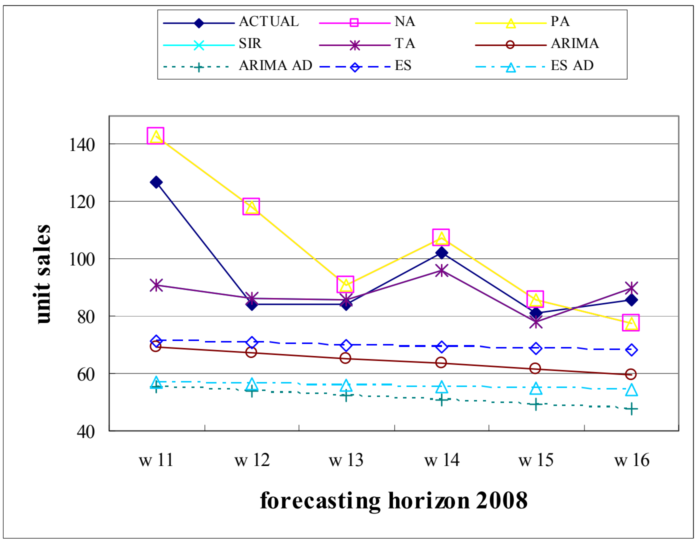

Figure C24. A comparison of various adjustment methods for item 24 in off period of 2008

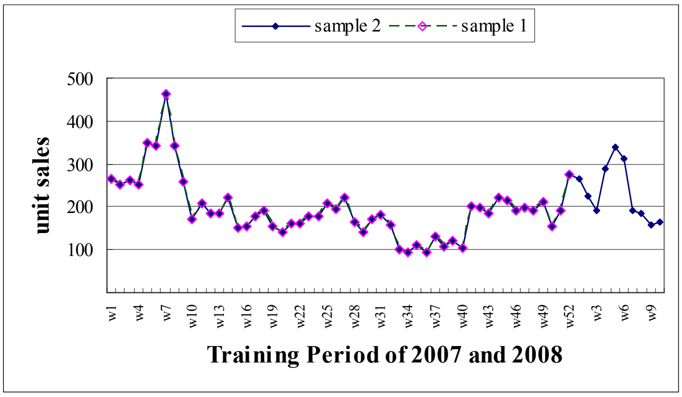

Figure A25. Sales data of item 25 in training period of 2007 and 2008

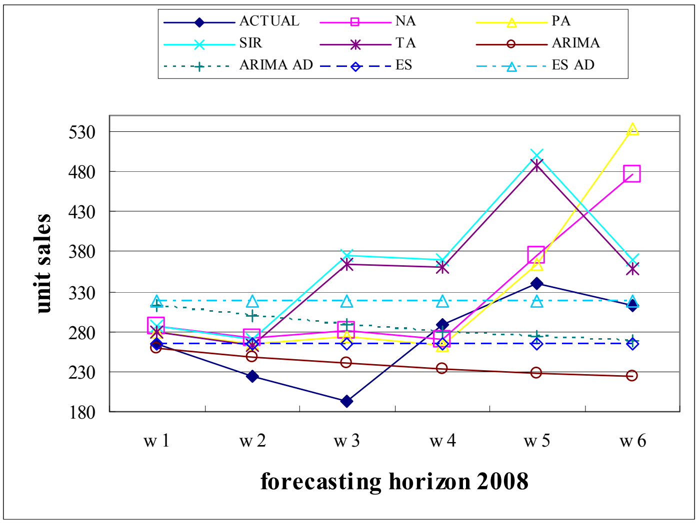

Figure B25. A comparison of various adjustment methods for item 25 in busy period of 2008

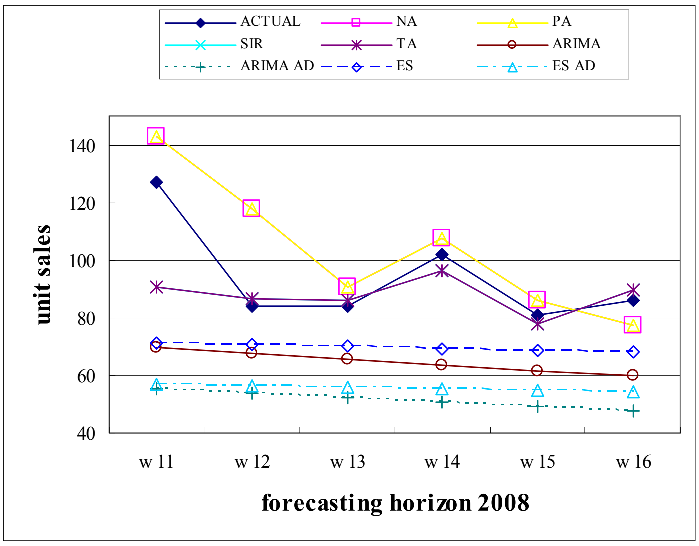

Figure C25. A comparison of various adjustment methods for item 25 in off period of 2008

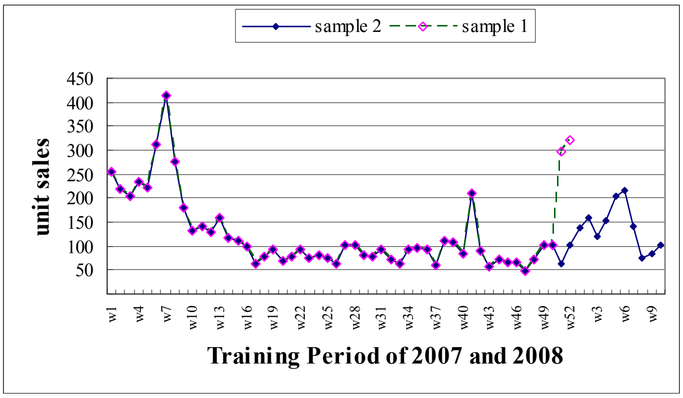

Figure A26. Sales data of item 26 in training period of 2007 and 2008

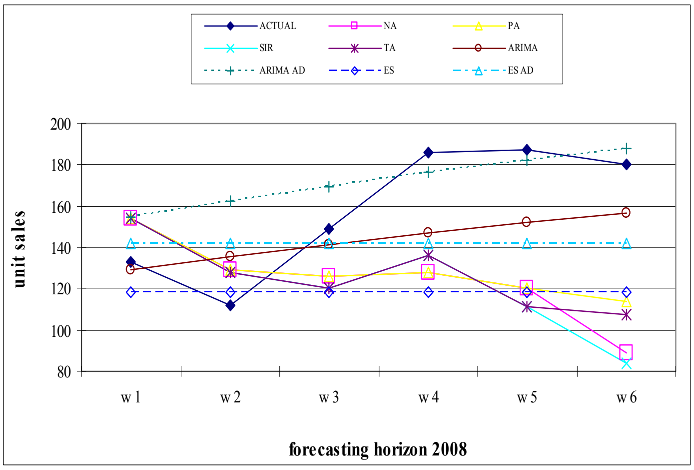

Figure B26. A comparison of various adjustment methods for item 26 in busy period of 2008

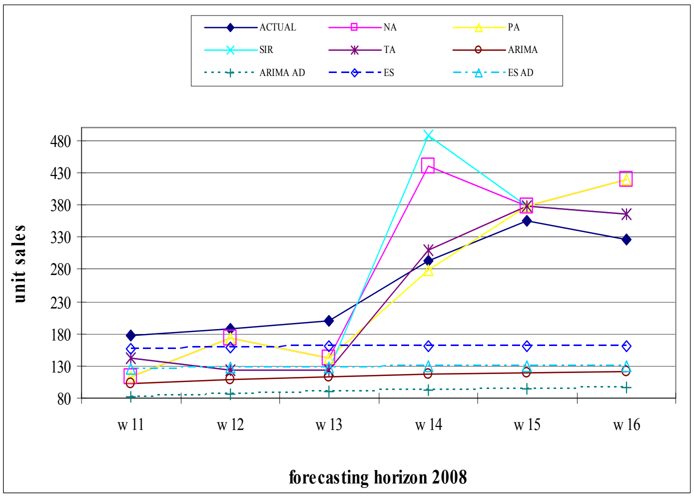

Figure C26. A comparison of various adjustment methods for item 26 in off period of 2008

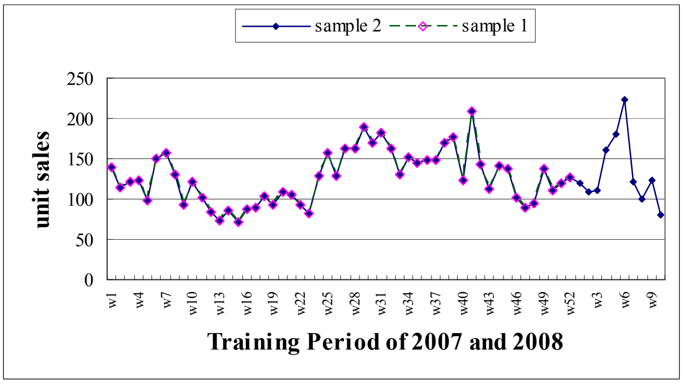

Figure A27. Sales data of item 27 in training period of 2007 and 2008

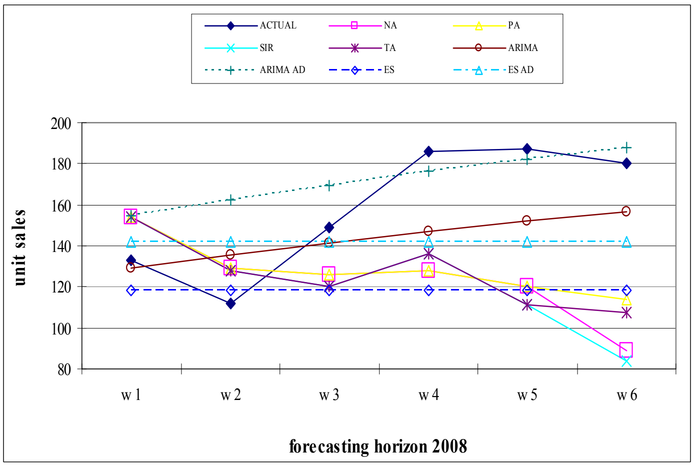

Figure B27. A comparison of various adjustment methods for item 27 in busy period of 2008

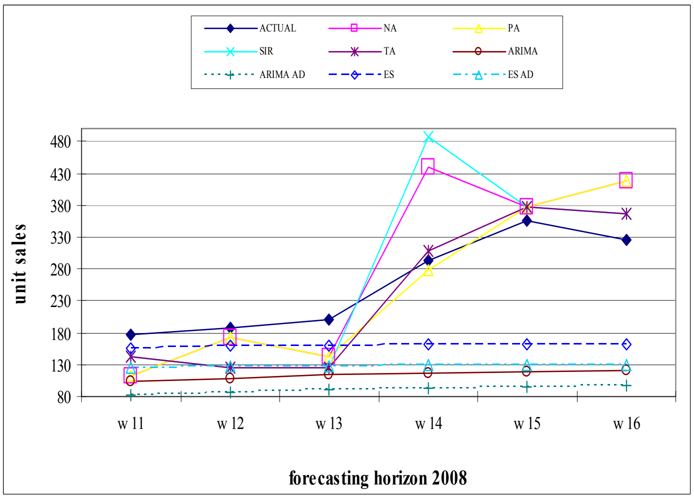

Figure C27. A comparison of various adjustment methods for item 27 in off period of 2008

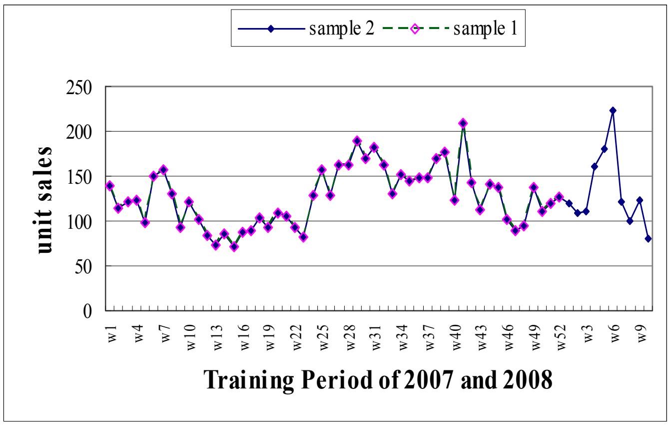

Figure A28. Sales data of item 28 in training period of 2007 and 2008

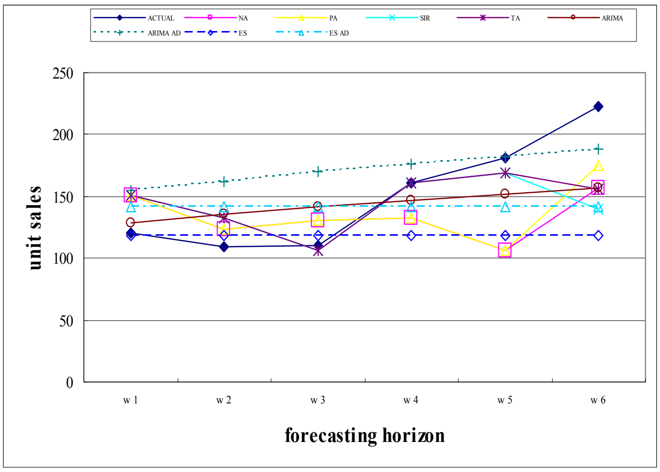

Figure B28. A comparison of various adjustment methods for item 28 in busy period of 2008

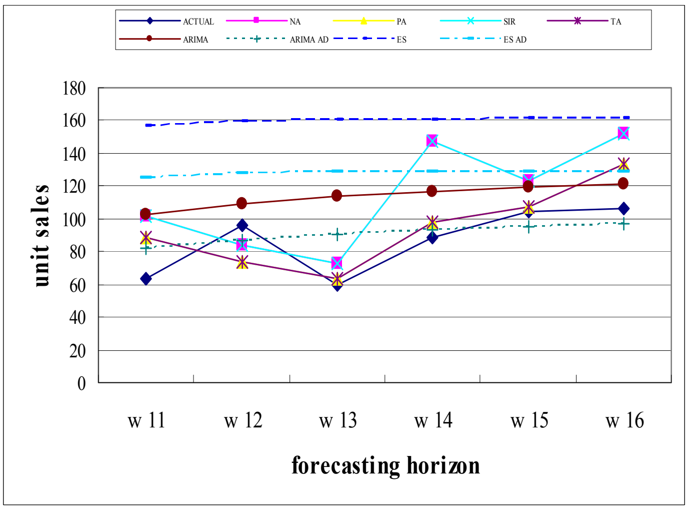

Figure C28. A comparison of various adjustment methods for item 28 in off period of 2008

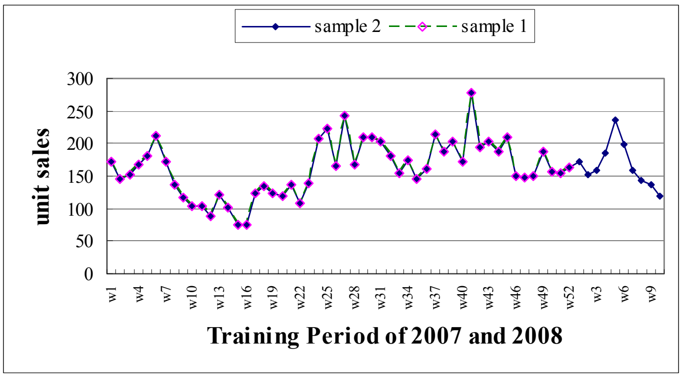

Figure A29. Sales data of item 29 in training period of 2007 and 2008

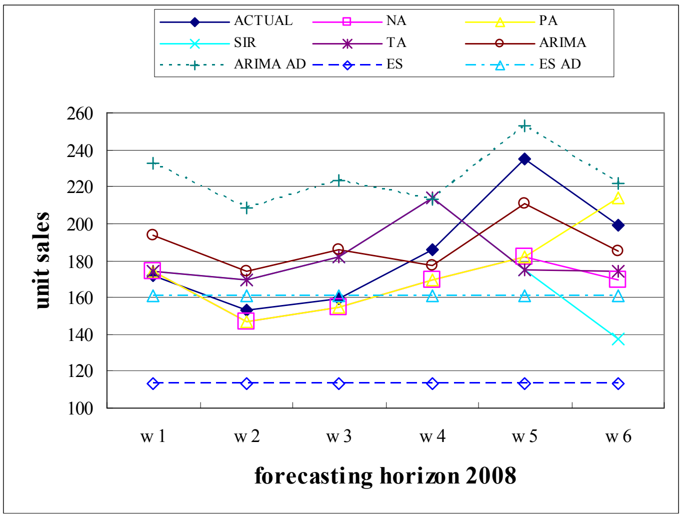

Figure B29. A comparison of various adjustment methods for item 29 in busy period of 2008

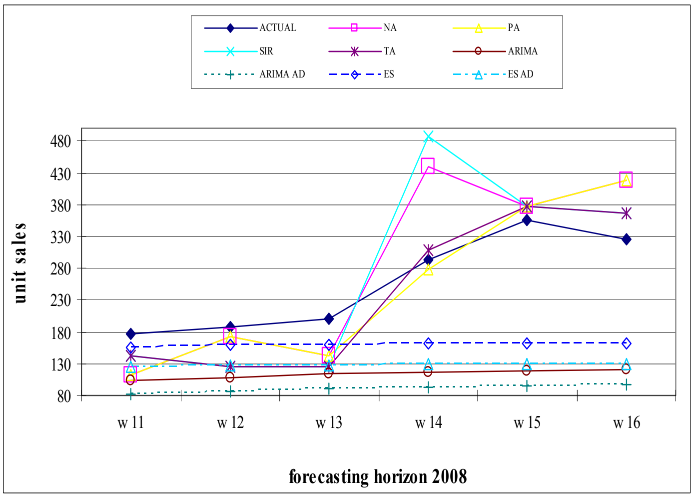

Figure C29. A comparison of various adjustment methods for item 29 in off period of 2008

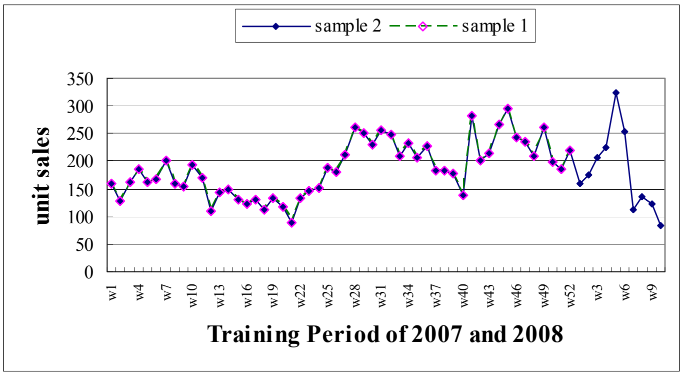

Figure A30. Sales data of item 30 in training period of 2007 and 2008

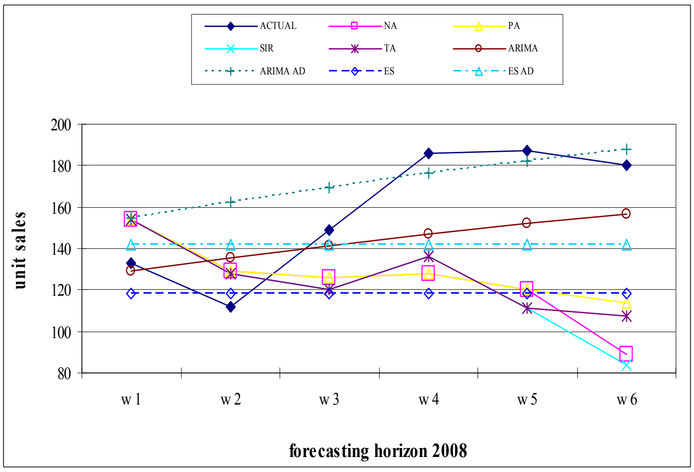

Figure B30. A comparison of various adjustment methods for item 30 in busy period of 2008

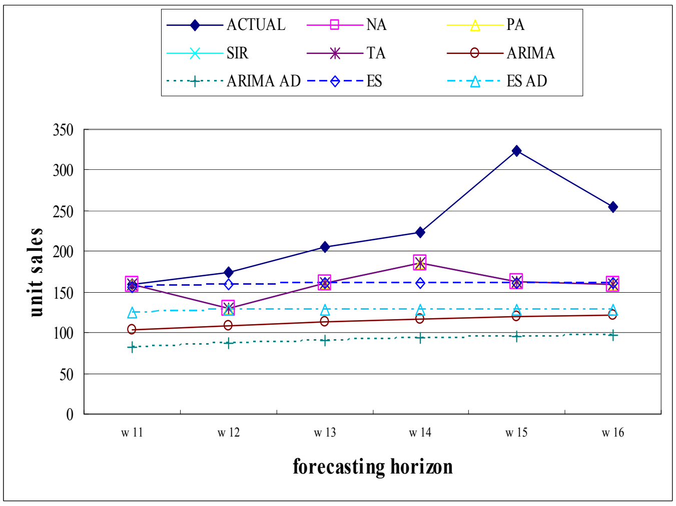

Figure C30. A comparison of various adjustment methods for item 30 in off period of 2008