

11-1-2010

# A GA-Based Sales Forecasting Model Incorporating Promotion Factors

Li-Chih Wang

*Tunghai University, Taichung, Taiwan ROC, wanglc@thu.edu.tw*

Chin-Lien Wang

*Ling Tung University, Taichung, Taiwan ROC, love123wang@gmail.com*

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Wang, Li-Chih and Wang, Chin-Lien (2010) "A GA-Based Sales Forecasting Model Incorporating Promotion Factors," *Journal of Modern Applied Statistical Methods*: Vol. 9 : Iss. 2 , Article 20.

DOI: 10.22237/jmasm/1288585140

## A GA-Based Sales Forecasting Model Incorporating Promotion Factors

Li-Chih Wang  
Tunghai University  
Taichung, Taiwan ROC

Chin-Lien Wang  
Ling Tung University  
Taichung, Taiwan ROC

---

Because promotions are critical factors highly related to product sales of consumer packaged goods (CPG) companies, predictors concerning sales forecast of CPG products must take promotions into consideration. Decomposition regression incorporating contextual factors offers a method for exploiting both reliability of statistical forecasting and flexibility of judgmental forecasting employing domain knowledge. However, it suffers from collinearity causing poor performance in variable identification and parameter estimation with traditional ordinary least square (OLS). Empirical research evidence shows that - in the case of collinearity - in variable identification, parameter estimation, and out of sample forecasting, genetic algorithms (GA) as an estimator outperform OLS consistently and significantly based on a log-linear regression model concerning weekly sales forecasting of CPG products from a manufacturer in both busy and off seasons.

Key words: Sales forecasting, genetic algorithm, ordinary least square, collinearity, variance influence factor.

---

### Introduction

Due to competition promotion has increasingly become a key factor of marketing in consumer packaged goods (CPG) industries because sales are highly related to promotion activities. To properly forecast unit sales of products for a particular company in the CPG industry, forecasters must take this contextual factor into account. Forecasts generated with most statistical models are consistent, but are usually devoid of the flexibility and comprehensiveness of contextual information. The lack of contextual information is exploited with judgmental forecasting, thus predictors and users of the forecasts are often tormented with the

issue of inconsistency due to bias. These issues are clearly pointed out by Sanders and Ritzman (1992), Armstrong and Collopy (1998), Webby, et al. (2001) and De Gooijer and Hyndman (2006) among many others.

Regression is a natural choice to connect both methods (Edmunson, 1990; Bunn & Wright, 1991; Armstrong, et al., 2005; Nikolopoulos, et al., 2006), because regression is able to incorporate critical contextual factors into the model and produces consistent results. In regression modeling, the classical ordinary least square (OLS) still is one of the most widely used estimators to identify significant factors and estimate parameters in linear regression (Draper & Smith, 1998; Rawlings, et al., 1998). However, it suffers from limitations posed by issues of outliers (Cook, 1977; Rawlings, et al., 1998; Meloun & Militky, 2001), 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 & Mason, 1977) in case of small size sample, regression models with highly correlated independent variables, and groups of dummy variables or sporadic

---

Li-Chih Wang is professor in the Department of Industrial Engineering and Enterprise information and the dean of the School of Engineering at Tonghai University, Taiwan, ROC. Email: wanglc@thu.edu.tw. Chin-Lien Wang is a lecturer in the department of Business Administration at Ling Tung University, Taiwan, ROC. Email: love123wang@gmail.com.

variables. As noted by Smith and Campbell (1980) it is ultimately caused by too little variation in predictor variables in the dataset which induce inflated variance of variable coefficients. Moreover, it usually causes many problems such as truly critical variables to become insignificant (Hendry, 2000) and incorrect parameter estimation in both sign and magnitude (Slinker, 1985), these problems will usually lead analysis and inference, as well as forecasting of the regression model to be out of track.

To address the issue of collinearity, an alternative parameter estimator called genetic algorithm (GA) is proposed; GA is an option to alleviate collinearity problems and obtain desired results with efficiency. This study begins with a log-linear regression model incorporating price and a group of non-price promotion related dummy variables (Kumar & Pereira, 1997; Heerde, et al., 2002a, 2002b). The model's effect parameters are assessed and decomposed with GA incorporating a fitness function of mean absolute percentage error (MAPE), which - without the square operation of errors. James and Stein (1961) exhibited an estimator which, under squared error loss, dominates the least squares estimator and, coupled with a realistic constraint on coefficient of variables, it is believed will - to some extent - avoid the issue of inflated influence of outliers and problems caused by collinearity in OLS.

This article proposes GA as an adequate alternative model estimator in regression modeling, particularly in situations of serious collinearity, through a comparative study of OLS and GA in in-sample parameters estimation and out-of-sample forecasting, respectively, with an empirical study on weekly unit sales forecasting of CPG products from a name brand manufacturer.

### Methodology

#### Formulation of a Regression Model

Equation (1) of the multiplicative regression model is motivated by Wittink et al.'s analytical models in a series of articles (Foekens, et al., 1999; Heerde, et al., 2002a, 2002b). Regression modeling uses a stepwise method called backward elimination (Draper & Smith, 1998), starting from the model

incorporating all critical factors considered, then removes insignificant variables one by one iteratively. The model can be formulated as

$$S_{it} = \lambda_{it} \left( P_{it} / \widehat{P}_i \right)^{\theta_{it}} \prod_{l=1}^n \mu_{lit}^{D_{lit}} \varepsilon_{it}, \forall t \in Q \quad (1)$$

where,

- $i$  denotes an item number,  $i = 1, 2, 3, \dots, I$ ;
- $t$  denotes specific number of period referenced;  
 $1 \leq t \leq T$ ,  $T$  is the total number of normal periods;
- $I$  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;
- $\lambda_{it}$  denotes the normal unit sales (base sale) of the item  $i$  in period  $t$  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;
- $\theta_{it}$  denotes the coefficient of price elasticity of item  $i$  during period  $t$  under a retailer;
- $D$  denotes an indicator parameter (or dummy variable) of non-price promotion mix;
- $D_{lit}$  is the  $l$ -th component of a vector of  $n$  indicator parameters of non-price promotion mix ( $D_{1it}, D_{2it}, \dots, D_{nit}$ ) of item  $i$  in period  $t$ .  $D_{lit} = 1$  denotes a promotion mix of type  $l$  arises, the default value of  $D_{lit} = 0$ ;
- $\mu_{lit}$  denotes the non-price promotion effect parameter (multiplier) of corresponding non-price promotion mix ( $D_{lit}$ ) of item  $i$  during normal period  $t$  under a retailer; and
- $\varepsilon_{it}$  denotes the residual error.

Taking the natural logarithm in both sides of (1) results in the following:

$$\ln S_{it} = \ln \lambda_{it} + \theta_{it} \ln \left( P_{it} / \hat{P}_i \right) + \sum_{l=1}^n D_{lit} \ln \mu_{it} + \varepsilon_{it};$$

$$\forall t \in Q \tag{2}$$

A nonlinear model such as (1) is transformed to a linear regression model (Carroll & Ruppert, 1988; Franses & McAleer, 1998), which is the underlying model to conduct model fitting and model checking in this study.

**Model Fitting: Parameter Estimation with GA**

The genetic algorithm (GA) is proposed as a regression estimator to identify critical variables and estimate coefficients of variables as opposed to the widely employed least square type of estimators in situations of small sample size or a model mainly composed of dummy variables and sporadic variables.

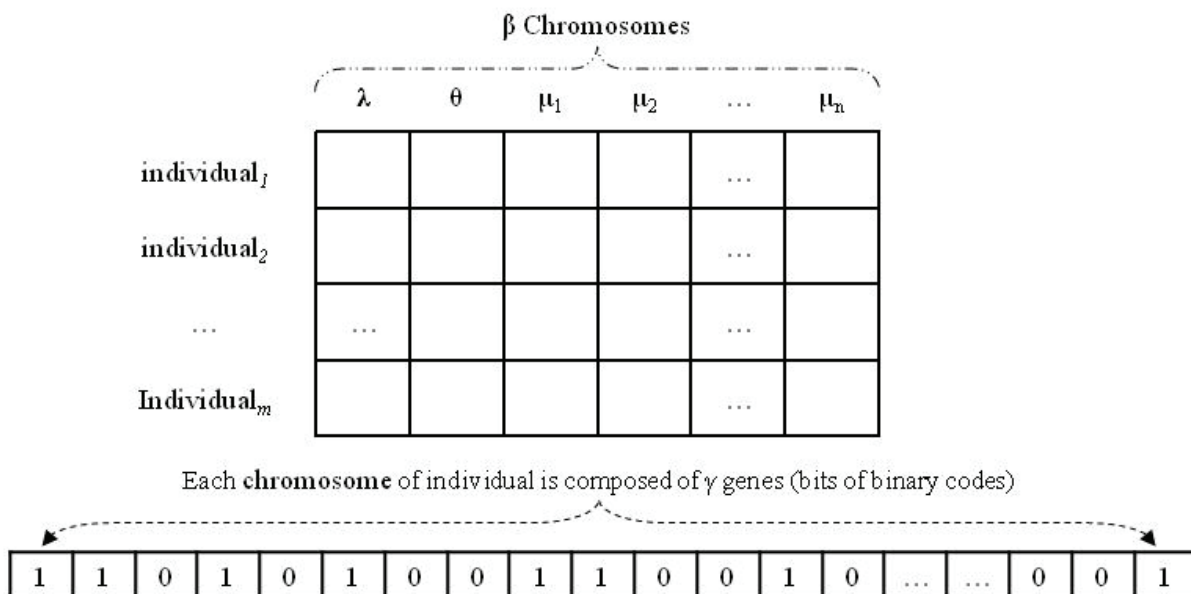
**Features of the Genetic Algorithm (GA)**

GA simulates Darwin’s biological evolution by selecting encoded individuals (solutions) in the population with higher fitness (via a fitness function) through stochastic crossover and mutation to generate a population of individuals (reproduction) more fitted to the environment (better solutions) from generation to generation (Holland, 1992; Goldberg, 1987,

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

The initial population is randomly created in the encoded form of a binary matrix, there exist  $m$  rows, each row of binary string in the matrix is an individual (solution) which encompasses  $\beta$  chromosomes, each chromosome represents a parameter and is composed of  $\gamma$  genes, each gene is represented by a binary code (See Figure 1). Each individual is evaluated by the fitness function as shown in Equation (3), in each generation, the best  $\alpha\%$  ( $1 \leq \alpha \leq 6$ ) of the population are kept as elites to the next generation, the remaining population is created by randomly selected pairs of individuals conducting a one-point crossover within each chromosome of such pairs to reproduce offspring, forming a random recombination of individuals’ ingredients of genes, to search for a new solution space and possibly a better solution. After this, a one-bit mutation is performed with a view to creating new pieces of

Figure 1: The Composition of Population Generated Randomly in GA



gene originally not possessed by members of the population through randomly selected genes within each individual; this occasional random change in genes could open the door to new possibilities of better solutions. Afterwards, each encoded individual in the population is decoded back to a string of real numbers of parameters and each individual is evaluated by the fitness function, this iterative process repeats until a termination condition is met.

Parameters such as crossover probability ( $P_c$ ) and mutation probability ( $P_m$ ) of GA are designed to vary with the number of generations processed or others, such as moving average percent of improvement (MAPI) in fitness function value within certain number of generations, to keep proper diversity of the population while retaining the convergence capability, to circumvent getting stuck too early at local solutions in its search process (Liu, et al, 2003; Pham & Karaboga, 1997).

Based on (2), the fitness function of GA may be formulated as

$$FV_i = MAPE_i = \frac{(\sum_{t=1}^T |\ln S_{it} - \ln \tilde{S}_{it}| / \ln S_{it})}{T}, \forall t \in Q \quad (3)$$

where the term  $|\ln S_{it} - \ln \tilde{S}_{it}|$  is the absolute value of difference between the natural logarithm of the actual sales volume ( $\ln S_{it}$ ) of the  $i^{\text{th}}$  item and natural logarithm of the estimated sales volume ( $\ln \tilde{S}_{it}$ ) of the same item in period  $t$ .  $T$  denotes the number of normal periods. The objective of GA 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. After model fitting, every effect parameter in (2) is derived in real value.

#### Model Checking

A regression diagnostics focused on normality and independence is performed to determine if critical assumptions of linear regression are violated, based on Equation (2). If these assumptions are severely violated,

particularly if collinearity arises among predictor variables, bias may be a serious issue in model fitting or in model specification.

The normality test is conducted through a one-sample Kolmogorov-Smirnov test (Lilliefors, 1967) and a Q-Q plot (Berilant, et al., 2005). An independence test in this study consists of two parts, namely, a multicollinearity test and an autocorrelation test. The former is performed via variance inflation factor (VIF), whereas the latter is performed via Durbin-Watson (D-W) test (Savin & White, 1977; Draper & Smith, 1998). 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

$$VIF_j = 1 / (1 - R_j^2), \quad j = n + 2. \quad (4)$$

where  $n$  denotes the number of types of non-price promotion mixes and  $R_j^2$  is the coefficient of determination from regression of the  $j^{\text{th}}$  predictor variable on the other predictor variables. As described in Theil (1971) and Berk (1977), estimated effect parameters can be directly proportional to  $VIF_j$  as the following equation:

$$s^2(\hat{\beta}_j) = VIF_j (\sigma^2 / (T-1)V_j^2) \quad (5)$$

where  $\hat{\beta}_j$  denotes the  $j^{\text{th}}$  effect parameters in equation (2),  $s^2(\hat{\beta}_j)$ ,  $\sigma^2$ , and  $V_j^2$  is the variance of  $\hat{\beta}_j$  and variance of regression errors, as well as the variance of the  $j^{\text{th}}$  predictor variable, respectively.  $T$  denotes the number of periods in the training period and can be perceived as sample size.

The D-W test focuses on testing whether any autocorrelation exists among the following series of regression error terms in equation (2) :

$\varepsilon_{it}, \varepsilon_{it-1}, \dots, \varepsilon_{i1}$ . The statistic can be formulated as

$$d = \sum_{t=2}^T (\varepsilon_{it} - \varepsilon_{it-1})^2 / \sum_{t=1}^T \varepsilon_{it}^2 \quad (6)$$

In general, as the serial correlation increases,  $d$  decreases.

The Re-composition of Variable Coefficients Estimated

The cycle length of CPG industry is about 52 weeks long, thus, let  $t' = t + 52$ , denoting the corresponding week to be forecasted in a new year. A naïve sales forecasting method considering cycle length to forecast unit sales of item  $i$  of period  $t'$  in a new year (see Williams, 1987), based on sales data of week  $t$  in the referenced year, would be

$$\ln \hat{S}_{it'} = \ln \eta_i + \ln \pi_{it} + \theta_{it} \ln \left( \frac{P_{it'}}{P_i} \right) + \sum_{l=1}^n D_{lit'} \ln \mu_{it'},$$

$$t' = t + 52, \forall t' \in Z \tag{7}$$

where  $\eta_i$  denotes the average normal sale of item  $i$  across referenced periods,  $\pi_{it}$  denotes the seasonal index of item  $i$  in period  $t$ , and  $Z$  denotes the set of periods to be forecasted. All parameters in equation (7) are derived either with GA or OLS. Let  $e_{1it'}$  denote the price effect multiplier of item  $i$  in forecasting period  $t'$  and  $e_{2it'}$  denote the effect multiplier of a non-price promotion mix. In each group of indicator parameters one condition at most will arise in each period, resulting in

$$\ln \hat{S}_{it'} = \ln \eta_i + \ln \pi_{it} + \ln e_{1it'} + \ln e_{2it'},$$

$$t' = t + 52, \forall t' \in Z \tag{8}$$

In its re-composed form, equation (8) can be used to forecast weekly unit sales. Parameters estimated through GA or OLS - based on observations in the training periods - can be recombined as in equation (8) to respond to expected promotional campaigns in the forecasting horizon (as specified in the promotion proposals) to perform out of sample forecasting with  $\ln \hat{S}_{it'}$  being transformed back to  $\hat{S}_{it'}$  in the following empirical study.

Empirical Study: Background

This study focuses on the forecast of weekly sales volume for several series of CPG products, manufactured by Company F, under retailer B. Company F is a leading manufacturer specialized in dehumidifier and deodorizer

products in Taiwan, and retailer B is an international outlet of DIY products. A sales data set of 10 items from 2007 and the first 4 months in 2008, aggregated from retailer B's outlets, coupled with price promotion, non-price promotion, and promotion proposals, are used to conduct the empirical study. The details of price rate and type of non-price promotion mix of these items are displayed in Tables 1a and 1 b. Each effect parameter is set to be constrained within a specific range in GA which was implemented in Matlab 6.5, for example, the price elasticity coefficient is set to be in the range of [-8, 0], while effect parameters of non-price promotion mixes are set to be between 1 and 5. However, the coefficients of predictor variables in OLS regression are estimated without any constraint in the statistical package SPSS 13.

Empirical Study: Experimental Design

In order to take both the busy season and off season into account and to have a proper assessment of the performance of both estimators, the forecasting horizon is designed to consist of two periods of equal duration, the first period includes the first 6 weeks of 2008 (one of the major busy seasons in that year) and the second period starts from the 11<sup>th</sup> week and ends at the 16<sup>th</sup> week of 2008 (one of the off seasons in that same year). The 10 product items manufactured by a name brand company of CPG products in Taiwan, in retailer B's outlets are the forecasting target in the empirical research.

To properly evaluate the performance of parameter estimation via GA and OLS as well as that of out-of-sample forecasting based on parameters derived from GA and OLS, respectively, particularly the consistency of performance, model fitting and checking is conducted with GA first and then with OLS consecutively, all based on the dataset of the entire year of 2007 as the first training period; this is a small period, thus, the training dataset in this period can be denoted as a small sample. The dataset for 2007 combined with the first 10 weeks of 2008 is the second training period is longer than the first, thus, the training dataset in forecast weekly unit sales of items of concern in the forecasting horizon.

WANG & WANG

Table 1a: Summary of Promotion Proposals for Year 2007 of Company F's Products under Retailer B

Item	Product Type	2007 Promotion Sessions, Content Denoted as ( $P_{it} / \hat{P}_i, D_i$ )*							
		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
1	Deodorizer	1, D <sub>4</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	89/99, D <sub>3</sub>	1, D <sub>2</sub>	85/99, D <sub>6</sub>	1, D <sub>1</sub>	79.5/99, D <sub>7</sub>
2	Deodorizer	59/65, D <sub>5</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	59/65, D <sub>3</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	59/65, D <sub>6</sub>	49.5/65, D <sub>7</sub>
3	Deodorizer	1, D <sub>4</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	119/138, D <sub>2</sub>	119/138, D <sub>2</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	99.5/138, D <sub>7</sub>
4	Dehumidifier	75/89, D <sub>5</sub>	75/89, D <sub>3</sub>	75/89, D <sub>2</sub>	75/89, D <sub>2</sub>	75/89, D <sub>2</sub>	75/89, D <sub>3</sub>	1, D <sub>1</sub>	75/89, D <sub>7</sub>
5	Dehumidifier	89/95, D <sub>5</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	89/95, D <sub>2</sub>	89/95, D <sub>3</sub>	89/95, D <sub>6</sub>	89/95, D <sub>7</sub>
6	Cleaner	90/109, D <sub>5</sub>	90/109, D <sub>1</sub>	90/109, D <sub>1</sub>	90/109, D <sub>2</sub>	89/109, D <sub>2</sub>	1, D <sub>6</sub>	90/109, D <sub>6</sub>	1, D <sub>6</sub>
7	Cleaner	85/89, D <sub>5</sub>	1, D <sub>1</sub>	85/89, D <sub>3</sub>	85/89, D <sub>3</sub>	85/89, D <sub>3</sub>	1, D <sub>6</sub>	1, D <sub>2</sub>	1, D <sub>6</sub>
8	Cleaner	195/219, D <sub>5</sub>	1, D <sub>1</sub>	195/219, D <sub>5</sub>	1, D <sub>1</sub>	195/219, D <sub>3</sub>	195/219, D <sub>6</sub>	195/218, D <sub>6</sub>	189/219, D <sub>7</sub>
9	Insect Pest	79/99, D <sub>5</sub>	79/99, D <sub>1</sub>	70/99, D <sub>1</sub>	70/99, D <sub>1</sub>	1, D <sub>1</sub>	1, D <sub>1</sub>	79/99, D <sub>6</sub>	1, D <sub>1</sub>
10	Insect Pest	52/65, D <sub>4</sub>	1, D <sub>1</sub>	52/65, D <sub>1</sub>	52/65, D <sub>3</sub>	52/65, D <sub>1</sub>	49/65, D <sub>6</sub>	49/65, D <sub>6</sub>	44.5/65, D <sub>7</sub>

\*Details of  $D_l, l = 1, 2, 3, 4, 5, 6, 7$ , can be checked in the Formulation of a Regression Model description

Table 1b: Summary of Promotion Proposals for Year 2008 of Company F's Products under Retailer B

Item	Product Type	2008 Promotion Sessions, Content Denoted as ( $P_{it} / \hat{P}_i, D_i$ )	
		12/27-2/12	2/14-4/1
1	Deodorizer	85/99, D <sub>4</sub>	1, D <sub>1</sub>
2	Deodorizer	55/65, D <sub>5</sub>	1, D <sub>1</sub>
3	Deodorizer	1, D <sub>4</sub>	1, D <sub>1</sub>
4	Dehumidifier	1, D <sub>5</sub>	1, D <sub>1</sub>
5	Dehumidifier	75/95, D <sub>5</sub>	1, D <sub>1</sub>
6	Cleaner	89/109, D <sub>5</sub>	1, D <sub>1</sub>
7	Cleaner	85/89, D <sub>5</sub>	1, D <sub>1</sub>
8	Cleaner	169/219, D <sub>5</sub>	1, D <sub>1</sub>
9	Insect Pest	1, D <sub>5</sub>	1, D <sub>1</sub>
10	Insect Pest	52/65, D <sub>4</sub>	1, D <sub>1</sub>

this period can be denoted as a large sample. Parameters derived from either estimator based both small sample and large sample are used to

### Results

#### Results of Model Fitting

The details of model fitting results are shown in Figure 4 and Tables A1-A4 in the Appendix. Tables A1-A2 are concerned with parameters estimated with GA on small sample and large sample, respectively, while Tables A3-A4 are concerned with parameters estimated with OLS on these two samples respectively. Most parameters derived from GA are consistent with expectations, such as, the effect parameters of  $\mu_1$  to  $\mu_3$  increase from 2.182 to 2.287 in busy-season periods and increase from 2.277 to 2.796 in off-season periods. This may be explained by more effort being made and more expenditure for promotions therein;  $\mu_5$  is larger than  $\mu_4$  because non-price promotion type 5 employs direct mail in addition to all aspects included in type 4,  $\mu_7$  is larger than  $\mu_6$  for the same reason. Effect parameters estimated by OLS also are roughly consistent with expectations; their magnitudes are much smaller than expected, however. For example, many are smaller than 1 which indicates a negative effect in promotion and seems unreasonable based on experience (see Tables A3-A4).

Nearly every intercept (normal sales) is inflated to the extent that it exceeds the unit sales of an item in a certain period and becomes difficult to explain based on daily life experience. However, the issue of difficult explanation for parameters derived (Mandel, 2007) is very common in least square type of estimators, including weighted least square and partial least square, in addition to OLS. Often critical variables are deleted from the model by OLS, for example, 3 variables are removed for item 8 based on small sample, price elasticity of item 5 is discarded in both samples, and in items 3 and 6 price elasticity coefficients are deleted in the model by OLS. These phenomena can lead to a dilemma of incapability to take advantage of certain domain knowledge or contextual information. Moreover, the price elasticity coefficients of item 2 from OLS in the large sample are positive (see Table A4) - a

phenomenon which goes against common sense, but the underlying reasons are now investigated.

#### A Comparative Analysis of Results in Model Checking via VIF and T-W Tests

The normality test, consisting of the one-sample Kolmogorov-Smirnov test and Q-Q plot, in which both GA and OLS passed the test with data from both small and large samples without difficulty. The independence test measures of VIF and the results of D-W tests, however, showed complex but interesting consequences in two training periods of different length via GA and OLS and warranted further investigation. As shown in Tables 2 and 3 the number displayed in each cell of these tables is the average VIF of a specific effect parameter of a certain item. The number in the cell in the right hand side column in the table is the mean of the average VIF for each item concerned.

Note that the mean of the average VIF in the first training period is much larger than that of the second training period, even though not every mean of the average VIFs in the first training period is necessarily bigger than its counterpart in the second training period. Some outliers arising in the first training period considerably increase the relevant measure. However, as Smith and Campbell (1980) note, although VIF can identify the source of inadequate parameter estimation, it cannot measure the amount of imprecision.

Because the main difference of the two training datasets is the sample size, one is 47 (5 cases are discarded as outliers in mixed periods which include two different kinds of promotions in a single week), whereas the other one is 56 (6 cases are discarded). The large sample seems to enable the predictor variables to have more changes in values within the dataset to alleviate the collinearity issue arising in the small sample based model. For example, as shown in Tables 2-3, the mean of average VIF reduces from 6.196 to 3.286 and the standard deviation reduces from 7.662 to 1.242 as the sample size increases from 47 to 56.



Table 2: The Results of Average VIF of Each Predictor Variable for Each Item Based on Small Sample

Item	Average VIF								
	Price Elasticity $\theta$	Pro-Mix $\mu_1$	Pro-Mix $\mu_2$	Pro-Mix $\mu_3$	Pro-Mix $\mu_4$	Pro-Mix $\mu_5$	Pro-Mix $\mu_6$	Pro-Mix $\mu_7$	Mean
1	1.523	20.139	15.038	19.374	--	23.929	28.459	19.104	18.224
2	1.293	2.357	2.230	2.275	2.265	--	3.785	3.708	2.559
3	1.072	1.160	1.143	2.134	2.922	--	--	1.046	1.580
4	1.345	2.801	--	2.382	3.111	--	2.868	4.636	2.857
5	7.902	2.547	1.290	7.737	--	7.664	1.257	--	4.733
6	1.376	1.483	1.900	1.605	--	1.605	1.312	1.436	1.531
7	1.265	3.993	4.992	--	--	6.996	3.724	--	4.194
8	1.410	1.521	1.655	--	--	1.958	1.673	--	1.643
9	1.191	2.099	1.972	--	2.278	--	2.278	--	1.964
10	1.229	40.567	21.733	27.930	--	21.486	23.077	--	22.670

As a result, the problem of deletion of critical predictor variables in model fitting with OLS seems to have been improved. For example, from item 2 to item 10, the number of discarded predictors decreases row by row with the exception of the item 9 row in both tables. In addition, the quality of parameter estimation conducted by OLS also seems to improve particularly for items with mean VIF greater than 10 (Craney & Surles, 2002), such as items 1 and 10 in the small sample, reduced to around 3 or less based on large sample (see Tables A3-A4 in the Appendix).

The much more serious issue of deleting predictor variables and the downgrade of parameter estimation quality owing to sample size change did not occur in model fitting with GA (see Tables A1-A2 in the Appendix), however. Compared with OLS, GA shows better and more consistent behavior in model fitting. The problems caused by the occurrence of collinearity among predictor variables of models based on a smaller dataset did not affect GA to a great degree in its parameter estimation. The reason may be attributed to the flexibility GA has in dealing with the dataset to comply with its purposes through the formulation of a fitness

function and the constraint of variable coefficients.

A comparative analysis was conducted as shown in Table 4 in which a D-W test was performed to check if any serious problem of serial correlation arising in the error terms of model fitting occurred. Roughly speaking, no big change concerning the condition of autocorrelation among sequential series of errors created in model fitting with GA based on different size of samples was observed. The number of cases rejected in the D-W test is 3 in the small sample, while in the large sample the number increases to 4. There exists an obvious change in the results of model fitting with OLS in this regard, the number of null hypotheses,  $H_0$ , rejected in the small sample is 6 out of 10, for large sample the number of rejected test cases reduces to 3. Apparently, for the small sample the condition of autocorrelation of regression errors created by OLS is more serious than that based on the large sample. However, regression errors created by GA did not show same kind of change between the small sample and large sample.

GA-BASED SALES FORECASTING MODEL INCORPORATING PROMOTION FACTORS

Table 3: The Results of Average VIF of Each Item

Item	Average VIF								
	Price Elasticity $\theta$	Pro-Mix $\mu_1$	Pro-Mix $\mu_2$	Pro-Mix $\mu_3$	Pro-Mix $\mu_4$	Pro-Mix $\mu_5$	Pro-Mix $\mu_6$	Pro-Mix $\mu_7$	Mean
1	1.337	4.413	2.896	2.544	--	3.775	4.863	3.091	3.274
2	1.269	1.907	1.923	2.038	1.685	--	3.692	3/211	2.086
3	1.326	7.514	8.089	6.148	6.295	--	--	1.882	5.209
4	1.322	2.906	--	2.694	2.969	--	3.362	5.597	3.142
5	8.327	2.005	1.349	8.273	1.956	8.208	1.313	--	4.490
6	1.305	3.007	2.913	2.646	2.913	2.646	5.488	2.434	2.919
7	1.271	5.149	6.260	--	--	7.917	4.522	--	5.024
8	1.396	1.538	1.796	--	--	1.690	1.882	--	1.660
9	1.203	2.369	2.253	--	2.151	--	2.635	--	2.122
10	1.217	3.456	3.076	3.836	--	2.927	2.820	3.197	2.933

Table 4: Results of the Durbin Watson Test for GA and OLS Respectively in Two Data Samples

Item	Durbin-Watson Test							
	GA				OLS			
	Small Sample		Large Sample		Small Sample		Large Sample	
d	Test Result	d	Test Result	d	Test Result	d	Test Result	
1	1.048	Inconclusive	1.030	Reject $H_0$	1.888	Reject $H_0$	1.085	Reject $H_0$
2	1.579	Inconclusive	1.282	Inconclusive	1.888	Reject $H_0$	1.576	Inconclusive
3	0.922	Reject $H_0$	0.797	Reject $H_0$	1.026	Reject $H_0$	1.097	Reject $H_0$
4	1.544	Inconclusive	1.106	Reject $H_0$	1.625	Reject $H_0$	1.225	Inconclusive
5	1.799	Reject $H_0$	1.616	Inconclusive	1.591	Reject $H_0$	1.350	Inconclusive
6	1.247	Inconclusive	1.117	Inconclusive	1.333	Inconclusive	1.223	Inconclusive
7	1.375	Inconclusive	1.744	Reject $H_0$	1.349	Inconclusive	1.689	Reject $H_0$
8	1.385	Inconclusive	1.367	Inconclusive	1.517	Inconclusive	1.429	Inconclusive
9	1.664	Reject $H_0$	1.460	Inconclusive	1.910	Reject $H_0$	1.910	Inconclusive
10	1.237	Inconclusive	1.286	Inconclusive	1.489	Inconclusive	1.508	Inconclusive

Figure 2: Comparative Forecasting Performance Based on Parameters Generated with GA and OLS on Small Sample

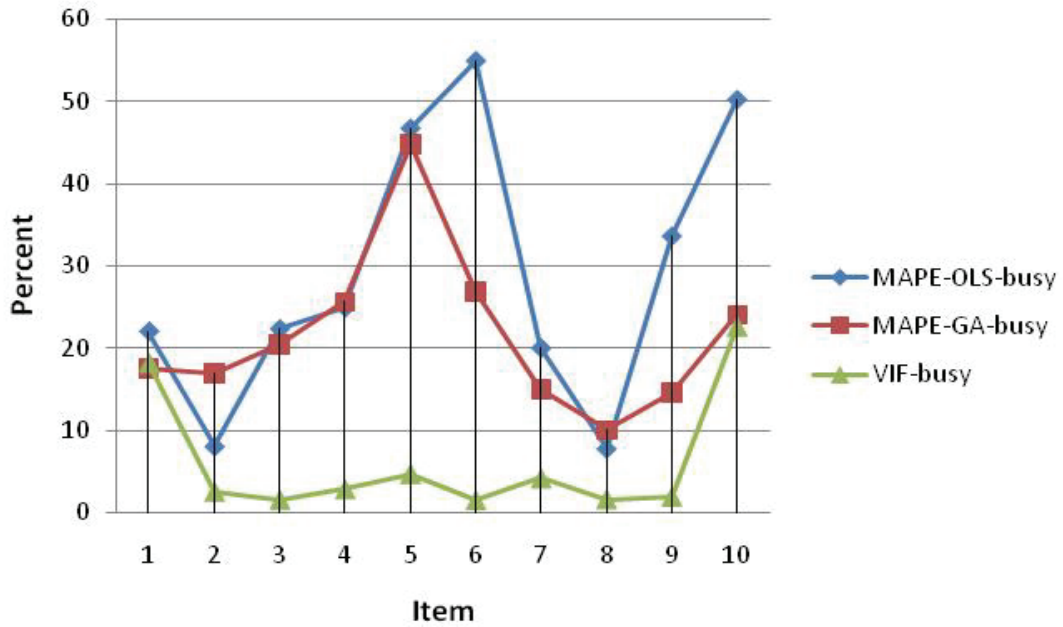


Table 5: Comparison of the Accuracy of Forecasting Based on Parameters Derived from GA and OLS

Item	MAPEs			
	Busy Season		Off Season	
	GA	OLS	GA	OLS
1	14.17%	22.09%	17.58%	26.53%
2	8.10%	8.10%	16.91%	20.94%
3	22.21%	22.38%	20.48%	20.68%
4	24.93%	24.93%	25.63%	26.89%
5	11.40%	46.67%	44.72%	47.58%
6	23.00%	54.86%	26.85%	37.33%
7	19.67%	20.05%	15.04%	43.75%
8	22.42%	7.81%	10.03%	17.21%
9	33.60%	33.60%	14.60%	14.50%
10	28.33%	50.17%	24.07%	27.14%
AVG	20.78%	29.07%	21.59%	28.26%

### Predictor Variable for Each Item Based on Large Sample

The model, estimated with two different methods based on two different sized samples is shown in Table 5. The superiority of the forecasting performance of GA over that of its counterpart is obvious: for forecasts in the busy season with parameters estimated from a small sample, except 4 cases, in which 3 cases are ties, only in one case did forecasting based on effect parameters derived from GA lose its ground to forecasting based on parameters generated with OLS. For forecasts in the off season the margin widens, 9 out of 10 items with parameters estimated via GA have an edge over those assessed by OLS, in terms of MAPE, in both seasons (see Table 5).

In addition, a paired-samples t test was conducted between MAPEs of forecasting based on parameters derived from GA and OLS on small samples  $t = -1.629$  at the  $\alpha = 0.10$  significance level (1-tailed) and the critical value is 1.383. Thus, the null hypothesis that, on average, the MAPE of OLS is smaller than or equal to that of GA is rejected and the alternative hypothesis that, on average, the MAPE of OLS is greater than that of GA is supported. The same paired sample t test results in a  $t = -2.459$ , rejects  $H_0$  and supports that, on average, the MAPE of OLS is greater than that of GA for a large sample at the 0.90 confidence level.

To further evaluate the effect of collinearity among predictor variables on the forecasting performance in either the busy or off season, two figures illustrate how and to what extent MAPE forecasting based on parameters derived from GA and OLS responds to the measure of VIF. For forecasting based on parameters derived from a small sample, on average, both GA and OLS show insignificant results between the MAPE of forecasts and the average VIF, 0.271 and 0.316, respectively, in the Pearson correlation test (2-tailed) with  $\alpha = 0.05$  level (see Figure 2 and Table 6). Conversely, using the same test, forecasting GA A paired t test (1-tailed) was performed, with  $\alpha = 0.05$ , between parameters estimated with GA on the small sample and large sample,

the t value =  $-0.547$  is greater than the critical value of  $-1.895$ , so it does not reject the

$H_0$  that, on average, parameters estimated with GA based on small sample are less than or equal to parameters estimated with GA based on large sample. Conversely, a paired t test, with the same  $\alpha=0.05$  between parameters estimated with OLS on the small and large sample results in  $t = 7.551$ , which is much greater than the critical value 1.895, thus it may be concluded at 95% confidence that, on average, parameters estimated with OLS on a small sample are greater than that for a large sample. Based on the above information, model parameters estimated with GA appear more stable than parameters and OLS based on parameters generated from large sample shows a significant result; the correlation coefficients are 0.618 and 0.649, respectively, (see Figure 3 and Table 6).

No significant difference exists between the performance of forecasting based on parameters derived on large or small sample of data for either GA or OLS. In sum, collinearity makes regression modeling with OLS more sensitive to a change in sample size so that the correlation between VIF and MAPE becomes less obvious in a small sample. Because low VIF is a necessary condition for good forecasting performance (Williams, 1987), a change from small to large sample does not create a significant difference in forecasting performance in terms of MAPE regardless of whether GA or OLS is used as estimator of model parameters.

### Conclusion

If a regression model is based on a limited size sample or if the variation of values in the dataset pertain to a specific critical variable that is too small, then the issue of collinearity will arise and make the model very sensitive to the sample size change and may negatively and seriously affect proper variable identification and variable coefficient assessment. Under such a situation, any analysis, inference or forecast based on the parameters of the model can be questionable. An alternative estimator, the genetic algorithm (GA), can - with proper formulation in fitness function and realistic constraints regarding coefficients of critical variables - have better and more consistent performance in both critical variable identification and variable coefficient estimation,

Figure 3: Comparative Forecasting Performance Based on Parameters Generated with GA and OLS on Large Sample

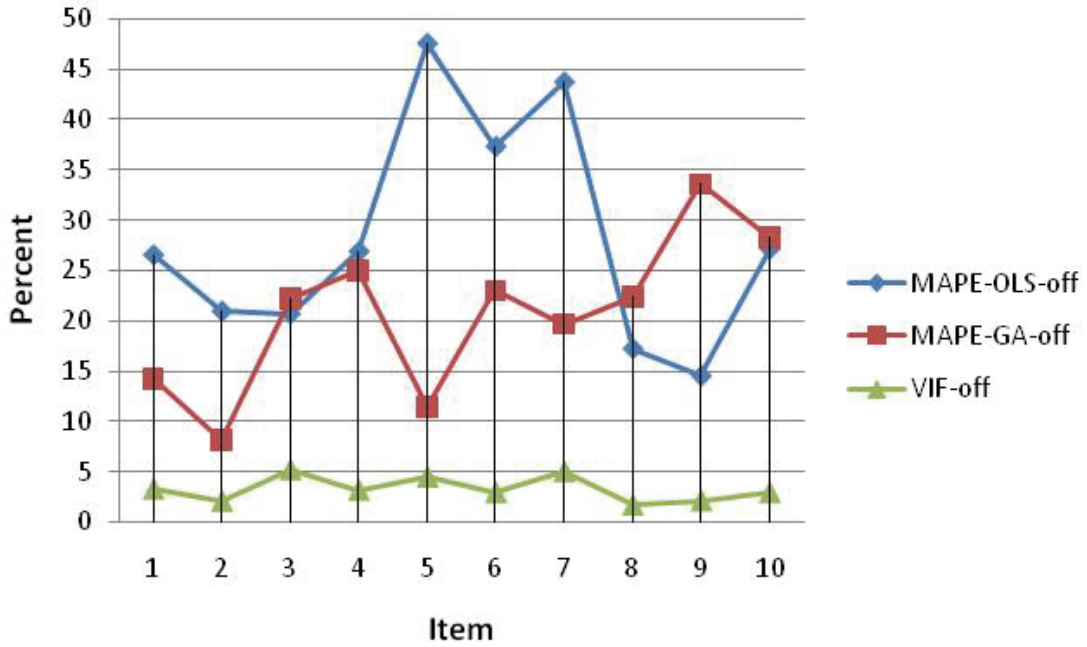
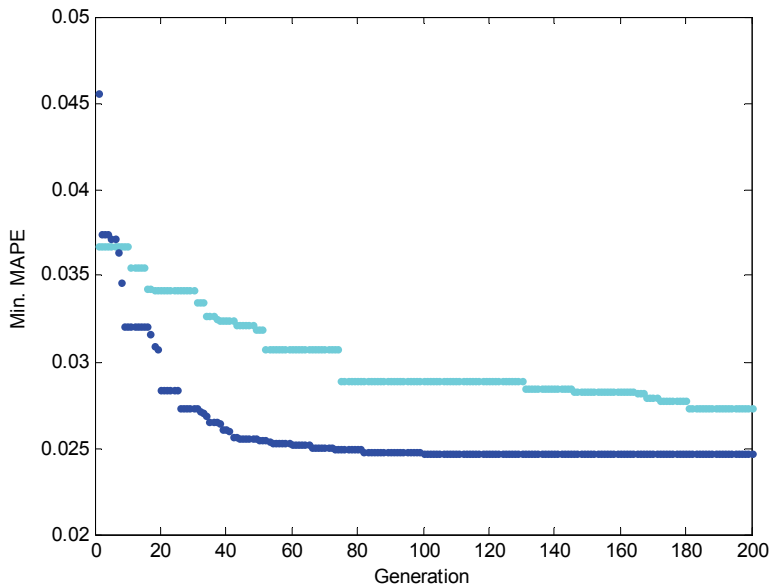


Table 6: Pearson Correlation Test Between VIF and MAPE

Estimator	Small Sample		Large Sample	
	Pearson Correlation	Significance (2 tailed) 0.05	Pearson Correlation	Significance (2 tailed) 0.05
GA	0.271	0.448	0.618	0.057
OLS	0.316	0.373	0.649	0.042

Figure 4: Typical Convergence Process of GA in this Study Compared to Generic GA



which can be verified via a series of measures, charts and model checking tests.

Empirical results support the points presented in this article via weekly unit sales forecasting based on a log-linear regression model of 10 CPG products from a name brand manufacturer in Taiwan in both a busy and an off season. More in-depth and wider investigations will be of great help to generalize points made in this article and to increase the amount of supporting data for use of the GA.

#### References

- Armstrong, J. S., & Collopy, F. (1998). Integration of statistical methods and judgment for time series forecasting: principles from empirical research. In G. Wright & P. Goodwin (Eds.), *Forecasting with Judgment*. Chichester, England: John Wiley.
- Armstrong, J. 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.
- 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.
- Belsley, D.A., Kuh, E. and Welsch, R.E. (1980) *Regression Diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley & Sons.
- 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.
- Bunn, D., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods: Issues and analysis. *Management Science*, *37*, 501-518.
- Burk, K. N. (1977). Tolerance and condition in regression computations. *Journal of the American Statistical Association*, *72*, 863-866.
- Carroll, R. J., & Ruppert, D. (1988). *Transformation and Weighting in Regression*. New York, NY: Chapman and Hall.
- Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*, 15-18.
- Craney, T, A., & Surlles, J, G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering*, *14*(3), 391-403.
- De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, *22*, 443-473.
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. New York, NY: John Wiley & Sons, Inc.
- Edmundson, R. H. (1990). Decomposition: A strategy for judgmental forecasting. *Journal of Forecasting*, *9*, 305-314.
- Eiben, A. E., Hinterding, R., & Michalewicz, Z. (1999). Parameter control in evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, *3*(2), 124-141.
- Foekens, E. W., Leeßang, P. H., & Wittink, D. R. (1999). Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometrics*, *89*, 249-268.
- Franses, P. H., & McAleer, M. (1998). Testing for unit roots and non-linear transformations. *Journal of Times Series Analysis*, *19*, 147-164.
- Goldberg, D. E. (1987). Simple genetic algorithms and the minimal deceptive problem, in: L. Davis (Ed.), *Genetic Algorithms and Simulated Annealing*. New York, NY: Hyperion Books.
- Goldberg, D. E. (1989). *Genetic algorithms, in search, optimization & machine learning*. Boston, MA: Addison-Wesley.
- Gunst, R. F., & Mason, R. L., (1977). Advantages of examining multicollinearities in regression analysis. *Biometrics*, *33*, 249-260.
- Hendry, D. F. (2000). Econometric modelling. Lecture notes for the PhD course in econometric modelling and economic forecasting, Department of Economics, University of Oslo.
- Van Heerde, H. J., Leeßang, P. H., & Wittink, D. R. (2002a). How promotions work: Scan\*Pro-based evolutionary model building. *Schmalenbach Business Review*, *54*, 198-220.
- Van Heerde, H. J., Leeßang, P. H., & Wittink, D. R. (2002b). Flexible decomposition of price promotion effects using store-level scanner data. *Schmalenbach Business Review*, *54*, 198-220.

- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.
- James, W., & Stein, C. (1961). Estimation with Quadratic Loss. In *Proceedings of the fourth Berkeley symposium mathematical statistics and probability, Vol. 1*. Berkeley, CA: University of California Press.
- Kumar, V., & Pereira, A. (1997). Assessing the competitive impact of type, timing, frequency, and magnitude of retail promotions. *Journal of Business Research, 40*, 1-13.
- Lim, J. S., & O'Connor, M. (1996). Judgmental forecasting with time series and causal information. *International Journal of Forecasting, 12*, 139-153.
- 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.
- 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.
- Meloun, M., & Militky, J. (2001). Detection of single influential points in OLS regression model building. *Analytica Chimica Acta, 439*, 169-191.
- Mandel, Igor. (2007). *Statistical modeling and business expertise, or where is the truth?* Working paper, Advanced marketing models, New York, NY.
- 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.
- 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.
- Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied Regression Analysis-A Research Tool*. New York, NY: Springer-Verlag.
- Sanders, N., & Ritzman, L. P. (1992). The need for contextual and technical knowledge in judgmental forecasting. *Journal of Behavioral Decision Making, 5*, 39-52.
- Savin, N. E., & White, K. J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica, 45*, 1989-1996.
- 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.
- Slinker, B. K., & Glantz, S. A. (1985). Multiple regression for physiological data analysis: the problem of multi-collinearity. *American journal of Physiology, 249*, R1-R12.
- Smith, G., & Campbell, F. (1980). A critique of some ridge regression methods. *Journal of the American Statistical Association, 75(369)*, 74-81.
- Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *The American Statistician, 45*, 53-56.
- Theil, H. (1971). *Principles of econometrics*. New York, NY: John Wiley and Sons, Inc.
- Webby, R., O'Connor, M., & Lawrence, M. (2001). Judgmental time series forecasting with domain knowledge. In Armstrong, J. S. (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwall, MA: Kluwer Academic Publishers.
- Williams, T. M. (1987). Adaptive Holt-Winters forecasting. *Journal of Operational Research Society, 38 (6)*, 553-560.
- 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.

GA-BASED SALES FORECASTING MODEL INCORPORATING PROMOTION FACTORS

Appendix

Table A1: Parameters Estimated via GA on Small Sample

Item	MAPE	Price Elas. $\theta$	Normal Sales $\lambda$	Pro- Mix $\mu_1$	Pro- Mix $\mu_2$	Pro- Mix $\mu_3$	Pro- Mix $\mu_4$	Pro- Mix $\mu_5$	Pro- Mix $\mu_6$	Pro- Mix $\mu_7$
1	0.124	-0.002	50.074	1.857	2.878	3.235	2.456	--	2.393	2.815
2	0.121	-0.0006	55.81	1.935	2.876	3.672	3.000	--	2.795	3.619
3	0.101	-0.273	110.62	2.190	--	2.193	--	4.163	--	2.340
4	0.107	-0.498	55.73	2.369	3.227	3.800	2.889	--	3.266	4.070
5	0.136	-0.488	200.00	2.750	2.300	1.614	--	3.436	1.969	3.951
6	0.121	-2.113	180.50	2.050	--	1.632	--	3.259	--	1.781
7	0.126	-2.749	79.09	1.808	--	2.698	3.132	--	2.720	2.834
8	0.092	-0.398	73.982	2.514	2.442	--	--	3.595	2.242	--
9	0.131	-8.768	91.265	1.709	1.911	--	2.614	2.951	1.854	--
10	0.236	-4.185	22.223	2.640	2.125	2.653	--	2.879	2.255	2.915
Mean	0.129	-1.947	91.929	2.182	2.537	2.687	2.818	3.381	2.437	3.041

Table A2: Parameters Estimated via GA on Large Sample

Item	MAPE	Price Elas. $\theta$	Normal Sales $\lambda$	Pro- Mix $\mu_1$	Pro- Mix $\mu_2$	Pro- Mix $\mu_3$	Pro- Mix $\mu_4$	Pro- Mix $\mu_5$	Pro- Mix $\mu_6$	Pro- Mix $\mu_7$
1	0.140	-0.048	42.83	2.171	3.362	3.764	2.872	--	--	3.257
2	0.121	-0.094	58.54	2.033	2.648	3.450	2.897	--	2.610	3.380
3	0.128	-1.078	78.035	2.845	--	2.875	--	3.170	--	2.665
4	0.130	-0.681	51.906	2.503	3.974	2.650	2.711	--	3.505	3.858
5	0.146	-0.100	199.68	2.414	2.304	1.728	--	3.680	2.938	1.559
6	0.126	-0.020	174.62	2.120	--	2.264	--	4.50	--	2.459
7	0.131	-2.765	87.99	1.625	--	3.809	2.967	--	2.436	1.083
8	0.097	-0.357	78.152	2.354	2.330	--	--	3.549	2.138	--
9	0.134	-7.110	74.700	2.236	2.520	--	3.408	--	2.443	--
10	0.237	-5.632	23.904	2.468	1.654	1.825	--	2.263	1.582	2.238
Mean	0.139	-1.789	87.036	2.277	2.685	2.796	2.971	3.432	2.522	2.562



Appendix (continued)

Table A3: Parameters Estimated via OLS on Small Sample

Item	MAPE	Price Elas. $\theta$	Normal Sales $\lambda$	Pro- Mix $\mu_1$	Pro- Mix $\mu_2$	Pro- Mix $\mu_3$	Pro- Mix $\mu_4$	Pro- Mix $\mu_5$	Pro- Mix $\mu_6$	Pro- Mix $\mu_7$
1	0.109	-0.098	430.30	0.214	0.3336	0.744	0.298	--	--	1.341
2	0.109	0.049	108	1	1.47	1.848	1.538	--	1.877	1.986
3	0.103	--	237	1	0.718	1.109	1.73	--	--	1.47
4	0.110	-0.61	132.44	1	--	1.545	1.241	--	1.367	1.543
5	0.129	--	309.00	1.836	1.433	1.230	--	2.487	1.959	--
6	0.099	--	375.00	--	1.370	1.182	--	2.457	1.321	1.333
7	0.129	-2.350	141.00	--	1.865	--	--	0.852	1.617	--
8	0.127	-0.210	188.00	--	1.000	--	--	--	--	--
9	0.133	-7.695	160.00	--	1.140	--	1.844	--	1.166	--
10	0.236	-6.54	61.00	--	0.600	0.687	--	1.041	0.617	0.804
Mean	0.128	-1.945	214.174	1.010	1.103	1.192	1.330	1.709	1.418	1.413

Table A4: Parameters Estimated via OLS on Large Sample

Item	MAPE	Price Elas. $\theta$	Normal Sales $\lambda$	Pro- Mix $\mu_1$	Pro- Mix $\mu_2$	Pro- Mix $\mu_3$	Pro- Mix $\mu_4$	Pro- Mix $\mu_5$	Pro- Mix $\mu_6$	Pro- Mix $\mu_7$
1	0.144	-0.098	121.80	0.768	1.178	1.329	1.103	--	--	1.169
2	0.116	-2.187	194.46	0.575	0.788	0.997	0.875	--	--	1.040
3	0.137	-2.908	265	0.852	0.642	0.725	--	--	--	0.480
4	0.142	-0.61	181.00	0.709	--	1.131	1.024	--	--	1.129
5	0.145	--	578.77	0.84	0.765	0.656	0.983	1.328	--	--
6	0.133	-0.69	495.00	0.757	0.912	0.787	1.514	1.637	--	0.888
7	0.159	-2.48	223.87	0.63	1.174	--	--	0.530	--	--
8	0.102	-0.233	175.75	1.077	1.068	--	--	1.706	--	--
9	0.138	-6.900	192.82	0.858	0.978	--	1.406	--	--	--
10	0.241	-4.065	51.310	1.197	0.943	1.170	--	1.630	--	1.350
Mean	0.146	-2.241	247.978	0.826	0.939	0.971	1.151	1.366	--	1.009