The value of competitive information in forecasting FMCG retail product sales and the variable selection problem

Abstract

Sales forecasting at the UPC level is important for retailers to manage inventory . In this paper, we propose more effective methods to forecast retail UPC sales by incorporating competitive information including prices and promotions. The impact of these competitive marketing activities on the sales of the focal product has been extensively documented. However, competitive information has been surprisingly overlooked by previous studies in forecasting UPC sales, probably because of the problem of too many competitive explanatory variables . That is, each FMCG product category typically contains a large number of UPCs and is consequently associated with a large number of competitive explanatory variables. Under such a circumstance, time series models can easily become over-fitted and thus generate poor forecasting resu lts .

Our forecasting methods consist of two stages. In the first stage, we refine the competitive information. We identify the most relevant explanatory variables using variable selection methods , or alternat ively, pool information across all variables using factor analysis to construct a small number of diffusion indexes. In the second stage, we specify the Autoregressive Distributed Lag (AOL) model following a general to specific modelling strategy with the identified most relevant competitive explanatory variables and the constructed diffusion indexes.

We compare the forecasting performance of our proposed methods with the industrial practice method and the AOL model specified exclusively with the price and promotion information of the focal product. The results show that our proposed methods generate substantially more accurate forecasts across a range of product categories.

1 introduction

Grocery retailers have been struggling with stock-outs for years. Stock-outs cause a direct loss of potential sales and lead to dissat­ isfied customers. The stock-out of individual items not only has negative impact on their own sales but also on the sales of the whole product category (Kalyanam, Borle, & Boatwright, 2007). Recent studies show that customers whom we once believed to either purchase substitutes or delay purchases when their preferred products are out of stock are actually more likely to switch stores and never come back (Corsten & Gruen, 2003 ). To avoid the out-of-stock condition, retailers may deliberately in­ crease safety stock (i.e. to over-stock), which substantially reduces profit (Cooper, Baron, Levy, Swisher, & Gogos, 1999 ). Under such a circumstance, retailers face a dilemma: they need to balance the loss due to stock-outs and the cost of safety stocks. One of the solu­ tions to resolve the service and cost trade-off is to employ accurate forecasts for product sales at the UPC level 1 (Corsten & Gruen, 2003 ). In supply chain management, accurate forecasts mitigate the bullwhip effect and are critically important for Just-In-Time UIT) delivery (Kuo, 2001; Ouyang, 2007; Sodhi & Tang, 2011 ).

Forecasting retailer product sales at the UPC level is difficult. Product sales are driven by a large number of factors such as marketing activities including price reductions and promot ion s. Today's grocery retailers spend a large proportion of their market­ ing budget on price reductions and promotions due to more intense competition (Kamakura & Kang, 2007; Raju, 1995 ). Price reductions and promotions substantially boost the sales of the focal product but also cause brand switching and stockpiling, which amplifies the variation of the product sales and makes prod­ uct sales more difficult to forecast (Ailawadi, 2006). The sales of the focal product are also subject to the negative impact of the price reduction and the promotion of other competitive products, which exacerbates the forecasting problem (Kamakura & Kang, 2007; Struse, 1987; Walters, 1991; Walters & Rinne, 1986 ).

In this study, we investigate the value of competitive informa­ tion including the competitive price and the competitive promo­ tion in forecasting product sales at the UPC level for retailers . We focus on the competitive information for two reasons . First, com­ petitive marketing activities including prices and promotions of the competitive products are important driving factors of product sales. Many studies in the marketing literature have been devoted to identifying and estimating the effect of these competitive mar­ keting activities, though they do not explicitly consider the opera­ tional question facing the retailer of designing models to forecast product sales at the UPC level (we review these studies in more depth in the next section). Therefore, there is a great potential gen­ erating more accurate forecasts by incorporating competitive information at the UPC level. One of main contributions of this pa­ per is the way in which competitive information are incorporated in the model to produce forecasts of product sales at the UPC level. Second , competitive information has previously been used to forecast product sales at the brand level. For example, Curry, Divakar, Mathur, and Whiteman (1995) incorporated the price and the promotion of three competing brands into a Bayesian VAR model to forecast the sales of the focal brand in the canned soup category. Divakar, Ratchford, and Shankar (2005) also incor­ porated the price, display, and feature of the main competitor (e.g. Pepsi versus Coca) to forecast the sales of the focal brand. The superior performance of their proposed model at the brand level implies the value of competitive information in forecasting product sales. However, forecasting product sales at the UPC level is a different research problem compared to at the brand level. Retailers rely on the forecasts at the UPC level for inventory plan­ ning, and use the forecasts at the brand level to arrange marketing tactics and as a negotiation tool in fulfilling agreements with man­ ufacturers Qiang, Klein, & Pick, 1998; Nijs, Dekimpe, Steenkamps , & Hanssens, 2001; White, 1986 ). Technically, the data at the disag­ gregate UPC level contains more noise than at the brand level and exhibit distinct patte rns. It is well known that the impact of competitive prices and competitive promotions are not as strong as the impact of the price and promotion of the focal product (see Hoch, Kim, Montgomery, & Rossi, 1995 ). Thus it is possible for the overall impact of competitive prices and promotions to be submerged in the noise of the data. In addition , there is intra-brand competition at the UPC level. Therefore, a related novel contribu­ tion of our research is the identification of key factors that affect

product sales at the UPC level.

The remainder of this paper is organized as follows . In section two, we review previous studies, which enables us to situate the main contribution of our paper. In section three we explain the problem of too many explanatory variables when incorporating competitive information at the UPC level. In section four we present our methodology. Section five describes the data. Section six intro­ duces the models. Section seven demonstrates our experimental de­ sign. In section eight we present the results . In the last section we draw conclusions on the value of competitive information in UPC re­ tail forecasting (both when the focal product is being promoted and when it is not) and also discuss directions for future research .

2 related literature

In practice, many retailers use the base-times-lift approach to forecast product sales at the UPC level. The approach is a two-step

procedure which initially generates a baseline forecast with simple time series models and then makes adjustments for any incoming promotional events. The adjustments are estimated based on the lift effect of the most recent promotion, and also the judgements made by brand managers (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Fildes, Nikolopoulos, Crone, & Syntetos, 2008). These judgmental adjustments are expensive and potentially prone to systematic biases, and evidence shows that the forecasting accu­ racy of this approach is far from satisfactory (Cooper et al., 1999; Fildes et al., 2009, 2008; Franses & Legerstee , 2010). In the more re­ cent literature, some studies focus on how to make the adjustment more effectively . For example, a string of studies have been de­ voted to helping managers with their judgmental decisions for the lift effect (Goodwin, 2002; Lee, Goodwin, Fildes, Nik olopoulos, & Lawrence, 2007; Nikolopoulos, 2010). Cooper et al. (1999) devel­ oped a model-based forecasting system to estimate the adjustment for the incoming promotion . The system estimates the lift effect of promotional events using a regression style model with a large number of variables related to the of price, promotion, and store/ category specific historical information . The system was subse­ quently extended to contain information related to manufacturers and product categories (Cooper & Giuffrida, 2000; Trusov, Bodapati, & Cooper, 2006). Apart from these studies , (Taylor, 2007 ) proposed quantile regression method to forecast supermarket product sales but ignored promotional information.

Other studies have attempted to forecast product sales by directly taking into account the promotional information. Rinne and Geurts (1988) considered the forecasting performance as a part of their evaluation of promotional profitability. Their model omitted dynamic and competitive effects and offered no evidence on forecasting accuracy. Preston and Mercer (1990) examined a limited number of product categories and again developed static models without competitive effects and there was no model com­ parison exercise being conducted. Kuo (2001) proposed a fuzzy neural network model to forecast daily milk sales. The neural net­ work model is integrated with a genetic algorithm which learns fuzzy rules obtained from marketing experts about the promo­ tional effects. However, their research only assessed the perfor­ mance of their approach in terms of one error measure (MSE) for one product (i.e. 500 cl container of papaya milk). Aburto and Weber (2007) proposed a hybrid model to forecast product sales for a Chilean supermarket at the SKU level. They initially forecasted the product sales with a seasonal ARIMA model and then predicted the residual using a neural network model with the price and promotional information of the focal product. Also they only analyzed their method in terms of one error measure (MAPE) for one product (i.e. vegetable oil, 1L). Giir Ali, Sayln, van Woensel, and Fransoo (2009) proposed various machine learning models to forecast retailer product sales at the SKU level. The model includes the support vector regression (SVR) and the regression tree methods with different priori settings. Their models incorporate a range of explanatory variables including the statistical information of unit sales, price, and promotions. They evaluated their model with a fixed forecasting origin and one error measure (MAE).

All the studies above tend to overlook some potentially impor­ tant features of the product market. For example , those focusing on the promotional event adjustment tend to ignore carryover effect of prices and promotions . Some studies incorporate stock variables (e.g. the total price reduction and promotion for the previous month) ignore the time information since the previous price reduc­ tion and promotion (e.g. Cooper et al., 1999 ). Most importantly, these studies overlook the potential importance of the price reduc­ tion and the promotion of other competitive products, apart from Giir Ali et al. (2009) who included stock variables to represent the overall level of competition intensity .

740 T. Huang et al./ European Journal of Operational Research 237 (2014) 738-748

The importance of the competitive information on product sales has been well established by previous research (e.g. Dekimpe, Hanssens, & Silva-Risso, 1999; Foekens, Leeflang, & Wittink, 1992; Nijs et al., 2001; Van Heerde, Gupta, & Wittink, 2003; Van Heerde, Leeflang, & Wittink, 2000). A well-known example is the SCAN\* PRO model and its extensions which measure cross price elasticity at the brand level (Andrews, Currim, Leeflang, & Lim, 2008; Wittink, Addona, Hawkes, & Porter, 1988 ). Other commer­ cially available solutions include PROMOTIONSCAN and BRANDAID (see Lilien, Kotler, & Moorthy, 1992; Lilien & Rangaswamy, 2004). More recent studies have analysed the cross price elasticity for individual items for each store (Wedel & Zhang, 2004 ). The nega­ tive impact of competitive marketing activities can be further di­ vided into the cannibalization effect and the brand switching effect depending on whether the impact originates from the prod­ ucts of the same brand or from different brands ( Nijs et al., 2001). Competitive information has been used in forecasting product sales at the brand level. For example, Foekens, Leeflang, and Witt­ ink (1994) examined the impact of the data aggregation level in the geographic dimension (e.g. the single store level, the chain level, and the market level) on the forecasting performance of the SCAN\*PRO model. However, they did not compare the SCAN\* PRO model with other conventional forecasting models because in fact they only wanted to compare relative forecasting performance in order to identify the impact of the data aggregation level on the SCAN\*PRO model. Curry et al. (1995) proposed a Bayesian VAR model to forecast canned soup product sales at the brand level. The model included the sales, price, and advertisement of four competing brands as endogenous variables and feature and display variables as exogenous variables. The Bayesian VAR model outper­ formed conventional time series models including Box-Jenkins models and the MARMA model in terms of the Theil's U statist ic. Divakar et al. (2005) proposed a regression model to forecast bev­ erage sales for manufacturers at the brand level. The regression model contains the price and promotion of the main competitors

(e. g. Coca versus Peps i), and also includes varying parameters to

take into account the heterogeneity across different distribution channels.

Table 1 summarizes the studies which forecast FMCG product sales at the UPC/SKU level and the brand level respectively. Many of these studies suffer from the problem of limited evaluation exer­ cise, e.g. too few products, inappropriate errors measures, the fail­ ure to use a rolling origin, and a fixed lead time design (Tashman, 2000). As a consequence researchers and especially those who use these models for decision making remain unsure about both the appropriate econometric specification and the relative accuracy of alternative and new proposed models. These earlier studies by neglecting the dynamics of the market and competitive effects leave unresolved various methodological questions which we now discuss.

3 the problem of too many explanatory variables

Previous studies have used competitive information to forecast product sales at the brand level (e.g. Curry et al., 1995; Divakar et al., 2005; Foekens et al., 1994 ). They typically incorporate com­ petitive information from the main competitors since there are only a small number of brands in the product category. However, there can be a large number of competing products at the UPC le­ vel: a typical product category such as Soft Drinks may contain hundreds of items of different flavours, package sizes, and brands which are all competitors with each other because they satisfy similar customer needs and wants (Kotle r, 1997 ). Accordingly we will have hundreds of competitive explanatory variables . Under such a circumstance , when we incorporate competitive

information, we face the problem of too many explanatory variables (Martin & Kolassa, 2009). Time series models can easily get over-fitted and generate poor forecasts and in an extreme case cannot even be estimated because of more explanatory variables than observations. Therefore a mechanism is needed to identify, select, and refine the most relevant competitive explanatory vari­ ables (Castle, Doornik, & Hendry, 2008). In this paper we propose a forecasting method which incorporates competitive information in forecasting retailer product sales at the UPC level. Methodolog­ ically our research propose an effective forecasting method which solve the problem of too many explanatory variables , an issue of theoretical and practical significance in a world of 'big data' . More importantly our research offer an operational guidance to the retail forecaster as to how to produce more accurate forecasts as simply as possible.

Many studies in the statistical field have been devoted to deal with this problem by seeking for an optimal subset of variables ( Miller, 1990 ). The most popular method is probably stepwise selection. The method starts with a null model and adds explan­ atory variables, step-by-step. At each step, the variable with the most significant contribution to the fit of the model is considered for addition while those variables in the model are examined to identify the one with the least significant contribution which is then considered for removal. In each case a threshold is estab­ lished to determine whether or not the action takes place. The process is complete when no additional actions meet the thresholds.

The stepwise selection method has been criticized for being likely to retain irrelevant variables and miss relevant variables ( Flom & Cassell, 2007; Harrell, 2001). Tibshirani (1996) proposed the Least Absolute Shrinkage and Selection Operator (LASSO) selec­ tion procedure as an alternative. The procedure estimates a regres­ sion model including all the potential explanatory variables but with a constraint for the sum of the absolute values of all the parameter coefficients. For example,

where

is the vector of observations on the dependent variable  
 is the matrix of the explanatory variables

*u* is the identically distributed random error

is the vector of unknown parameters  
*N* is the number of parameters  
 is the shrinkage factor which equals to the sum of all the parameter coefficients.

where *Y* is the vector for the standadized dependent variable, *X* is the matrix of the standadized explanatory variables, *u* is the identi­ cally distributed random error, *fJ* is the vector of unknown parame­ ters, *N* is the number of parameters, and *r,0* is the shrinkage factor which is equal to the sum of all the parameter coefficients.

When the constraint is employed, some of the parameter coef­ ficients will tend to be zero, and their corresponding explanatory variables will be removed from the regression model. In the selec­ tion procedure, the shrinkage factor is determined by the Akaike Information Criterion (AIC).

Flom and Cassell (2007) compared the performance of LASSO with stepwise selection using simulation approach. Their results suggest that stepwise selection tends to miss relevant explanatory variable when sample size is small and also retain irrelevant explanatory variables, while LASSO has better performance . How­ ever, as stated in Efron, Hastie, Johnstone, and Tibshirani (2004) there is no perfect variable selection method simply because none of them utilize any domain knowledge .

Variable selection methods identify the most relevant compet­ itive explanatory variables and the performance of the resulting forecasting model relies exclusively on these variables. Alterna­ tively, we may pool information across all the competitive explan­ atory variables and condense them into a small set of estimated factors at an acceptable cost of information loss (Stock & Watson, 2002a, 2002b). Many studies in the macroeconomics literature used factor analysis to summarize variations among a large set of variables (e.g. Engle & Watson, 1981; Forni & Reichlin, 1996 ). In part icular, Stock and Watson (2002b) constructed a number of fac­ tors (named as "diffusion indexes") with factor analysis to measure the common movement in a set of macroeconomic variables, and then used them to forecast real economic activities such as price infl ation. Their "dynamic factor" model has the following form :

where

is an *N*-dimensional multiple time series of explanatory variables

is the matrix with common factors of latent diffusion indexes

is the t value of the dependent variable

is a vector of the lagged dependent variable

and are the vectors of the parameter coefficients

and are the errors which are assumed to be and uncorrelated with each other.

where *Xt* is an N-dimensional multiple time series of explanatory variables, *Ft* is the matrix with r common factors of latent diffusion indexes, *Yt* is the value of the dependent variable, *Qt* is the vector of

the lagged dependent variable, and*' co* are the vectors of the

*'F*

parameter coefficients , and *et* and *Et* are the errors which are as­

sumed to be iid and uncorrelated with each other.

In the model, the original N competitive explanatory variables, Xt , have been condensed into r diffusion indexes at a cost of infor­ mation loss (i.e. et ), Stock and Watson (2002b) found that much of the variation in a large number(>100) of macroeconomic time ser­

ies (i.e. 39% of the total variation) can be accounted for by only six diffusion indexes. Their proposed models with diffusion indexes outperform the benchmark autoregressive models and VAR mod­ els, and they found that the models with the best forecasting per­ formance contained no more than one or two diffusion in dexes

4 methodology

In this study, we incorporate competitive information to fore­ cast retailer product sales at the UPC level. Considering the impor­ tance of the problem of too many explanatory variables, we propose a forecasting method with two stages. In the first stage, we refine the competitive information we want to incorporate in the forecasting model. Specifically, we implement both the vari­ able selection method and also the factor analysi s. For the variable selection method, we apply both the stepwise selection and the LASSO selection procedure, and we take the explanatory variables selected by the two methods in combination. This reduces the pos­ sibility of missing important explanatory variables (i.e. under­ specification, which causes the parameters to be biased and the inference of the parameters to be invalid), thereby making the

specified model more robust, though with a cost of efficiency. For the factor analysis, we construct diffusion indexes based on the competitive prices and the competitive promotions separately. We choose the most representative factors (e.g. those with eigen­ values substantially larger than others) while keeping the number of factors as small as possible, following the findings by Stock and Watson (2002b) .2

In the second stage, we incorporate the refined competitive information into econometric forecasting models . In this study, we construct the Autoregressive Distributed Lag (ADL) model fol­ lowing a general-to-specific modelling strategy ( Hend ry, 1995 ). We choose the ADL model for several reasons. First , the ADL model has the advantage of taking into account the carryover effect of the price and promotional variables . Second , the general-to-specific modelling strategy ensures the parsimony and data congruence of the model. Third, the ADL model is transparent with a simple regression style model struct ure, which benefits the users (Fader & Hardie, 2005). It has good interpretability compared to "black box" machine learning approaches which can hardly be under­ stood by brand/category managers. Also, in the forecasting litera­ ture, the general-to-specific ADL model is one of the most popular time series forecasting model and has exhibited superior forecasting performance in other areas including manufacturer sales, to urism , and air passenger flows (see Albertson & Aylen, 2003; Fildes, Wei, & Ismail, 2011; Song & Witt, 2003). Specifically , we start with a general model assuming that it properly describes the salient features of the data generating process, and then sim­ plify the general model by seeking out valid parsimonious restric­ tions. The following example shows the general ADL model with the most relevant competitive explanatory variables identified by the stepwise selection and the LASSO selection procedure:

where

is the log sales of the focal product at week

is the log price of the focal product at week

is the promotional index of the focal product at week

is the log price of competitive product at week

is the promotional index of competitive product at week

and are the numbers of competitive price and promotional variables selected by the variable selection methods

is the four-week-dummy variable  
 is the dummy variable for the calendar event at week . When , the dummy variable represents the week of the calendar event, and the week before the event if . takes the values from 1 to 9 representing all the calendar events *[[1]](#footnote-1)*

are the parameters  
 is the error term and we assume

is the order of the lags[[2]](#footnote-2)

where ln(Yo,t) is the log sales of the focal product at week *t,* ln(Po ,t- j) is the log price of the focal product at week *t* - *j, Promotiono,t-i* is the promotional index of the focal product at week *t- j,* ln(P m.t- j) is the log price of competitive product m at week *t* - *j, Promotionn,ti-* is

the promotional index of competitive product *n* at week *t* - *j, M* is the number of competitive price variables selected by the variable selection methods, *N* is the number of competitive promotional variables selected by the variable selection methods, *Four\_week\_dummyd* is the dth four-week-dummy variable, *Calen­ darE ventc,t- v* is the dummy variable for the cth calendar event at week *t* - *v.* The dummy variable represents the week of the calen­ dar event when *v* = 0, and the week before the event if *v* = 1. c takes the values from 1 to 9 representing all the calendar events3 ; *IXj, /Joj,*

)'oj, */Jmj, Y nj , 0d, bc,v* are the parameters, *Et* is the error term and we assume *Et* iid(0 , er), and *L* is the order of the lags.4

The general ADL model will ideally pass all the misspecification tests (e.g. the *F-test,* the Breusch-Godfrey test for autocorrelation, and tests for heteroskedasticity and normality). The model may be estimated by OLS with the usual interpretations of the statistics whether or not the data series are stationary, because sufficient lags were included to remove any autocorrelation (although with some potential loss of efficiency) (Song & Witt, 2003). A well­ specified ADL model can then be simplified following the gen­ eral-to-specific strategy. For example, we first estimate the general ADL model and remove the explanatory variable with the highest p-value for the parameter restriction test. We then estimate the reduced model and re-conduct all the misspecification tests. If the reduced model passes all these tests, we move onto remove the variable with the highest p-value in the new estimation, pro­ vided that the previous variable has already been removed, and so forth. Otherwise we will add the variable back and repeat the process by removing the variable with the second highest p-value for the parameter restriction test. In the modelling process we also remove the variables with incorrect signs and those not economi­ cally significant (i.e. with very small parameter coefficients) to achieve parsimony. The final simplified ADL model must pass all the misspecification tests of the general ADL model. The model is estimated by OLS with robust estimators in the presence of heter­ oscedasticity. Analogously, the following example shows the general-to-specific ADL model with the diffusion indexes:

where

is the diffusion index of competitive prices at week .

is the diffusion index of competitive promotion at week .

*P* and *Q* are the numbers of initially retained diffusion indexes, and [[3]](#footnote-3)

5 the data

In this study we use the weekly data from *Dominick's Finer Foods,* a large U.S. retail chain in the Chicago area. The data is pub­ licly available from the University of Chicago website.5 An advan­ tage of using this dataset is that a large number of studies have been conducted based on this dataset and many of them focus on identifying and measuring the effectiveness of the marketing mix activities (e.g. Fok, Horvath, Paap, & Franses, 2006; Kamakura & Kang, 2007; Song & Chintagunta, 2006 ). However, perhaps surpris­ ingly given the importance of forecasting at the UPC lev el, none of the studies using this dataset focuses on evaluating the performance of forecasting models. The dataset contains product information at the UPC level including unit sales, price, and promotions for 399 weeks. There are three different types of promotions: *"Simple* price *reduction", "Bonus buy",* and *"Coupo ns"."Bonus buy"* is the dom­ inant type, which corresponds to over 75% of the all the promotional events: 24.5% of promotions are *"Simple price reduction";* only less than 0.5% of promotions are *"Coupons".* In this study, we use one sin­ gle variable to represent the presence of all the promotional activi­ ties. We aggregate the data across 83 stores using constant weights based on the percentage of All Commodity Volume (ACV) of each store (see Hoch et al., 1995; Pauwels & Srinivasan, 2004 )6. In this study, we conduct our evaluat ion based on 122 products from 6 diverse product categories including *Bottled Juice, Soft Drinks, Bath Soap, Front-End-Candies , Frozen juice, and Bathroom Tissue .7*

Table 2 summarizes the characteristics of the data series for the 122 products during a time period of 200 weeks. First , it summa­ rizes the promotional intensi ty. For example , we choose 34 prod­ ucts from the *Bottled Juice* category. On average, these 34 products are being sold on promotion for 42 weeks during the 200weeks considered (i.e. an intensity of 0.21, with a standard deviation of 0.09). Second, it summarizes the average promotional index value. For example , the average promotional index value for the *Bottled juice* category is as high as 0.78 (with a standard devi­ ation of 0.05) , which indicates that the products in this category tend to be promoted simultaneously across all the selected stores. Third, Table 2 summarizes the lift effect of the promot ions . Take the *Bottle juice* category as an example, the promotions in this cat­ egory increase the sales of focal product by 169% on average com­ pared to the baseline predicted sales assuming there were no promot ion. Finally, Table 2 exhibits the average ratio of standard deviation *versus* mean for both sales and price of the products in each category. Among these product categories, *Bath Soap* and *Front-end-Candies* have the least variations for their product sales and price, and they also have the least intensive promotions . In contrast , *Soft Drinks* and *Bathroom Tissues* are heavily promoted and exhibit highly variations for the sales and price of their prod­ ucts . Therefore, our study covers data of a wide range of sales and promotional conditions.

Fig. 1 is an example for one product in the *Bottled juice* category (i.e. Tree Top Apple Juice 4802). The figure exhibits its unit sales , price (in US dollar) , calendar events, and promotional periods which are highlighted in darker bars. The length of the darker bars indicates the value of the promotional index which is between 0 and 1. The price and promotional index are both aggregated across multiple stores based on the percentage of ACV of each store. We applied the Augmented Dickey-Fuller test to investigate the stationarity of the data series for all the 122 products, and we find that most data series are station ary.8

6 the benchmark models

In this study we consider the following benchmark models : (1) the robust simple exponential smoothing (SES) model which focuses exclusively on the pattern of previous product sales; and

(2) the industrial base-times-lift approach which first produces baseline forecasts and then makes adjustments for any incoming promotional event. we implement the base-times-lift approach following Gilr Ali et al. (2009) :

where Mr is the baseline forecast for week *t* generated by a simple exponential smoothing model. Sr\_1 is the actual sales value in the previous week when the focal product was not on promotion. *a* is the parameter which is estimated by minimizing the mean squared error in the estimation period. The adjustment is calculated as the increased sales from the most recent promotion of the focal prod­ uct. In this study, we use aggregated data across multiple stores, thus the effects of promotions are represented by promotional indexes instead of promotional dummies. For example, if the most recent promotion has a promotional index value of 0.6 and we

8 We find 110 out of 122 data series for unit sales, 106 out of 122 data series for price, and all the data series for promotional indexes as stationa ry.

consider the "lift" effect as L. Then the adjustment for the forthcom­ ing promotion with an index value of 0.9 will be {0.9/ 0.6) \* L = 1.SL. In this study we propose two forecasting methods which both capture the effect of competitive information but in distinct ways. The first method is the general-to-specific ADL model with the most relevant competitive explanatory variables identified by the variable selection methods (i.e. the ADL model). The second is the general-to-specific ADL model with the diffusion indexes con­

structed using factor analysis (i.e. the ADL-DI model). We include the competitive price and promotion variables for most products of each product category.9 To understand the value of the competi­ tive information, we also include the general-to-specific ADL model which is constructed exclusively with the price and promotional information of the focal product (i.e. the ADL-own model).

7. experimental design

All the studies we have identified which forecast product sales were conducted with a single fixed forecasting origins (e.g.[Ali et al., 2009](#_ENREF_4); [Cooper et al., 1999](#_ENREF_15); [Divakar et al., 2005](#_ENREF_23)). However, evaluation results based on single forecast origins can be unreliable when the forecasting results are sensitive to both randomness and possible systematic business cycle effects ([Fildes, 1992](#_ENREF_26)). In this study, we evaluate the performance of our models with 70 rolling forecast origins, which partially controls for the effect of any specific effects arising from a particular origin. Forecast horizon should also be fixed in any forecast comparisons. We first estimate the models with a moving window of 120 weeks and forecast one to weeks ahead. The forecast horizons were chosen to take into account typical ordering and planning periods, and we set to be 1, 4, and 12. We then move the estimation window forward week by week throughout the remaining sample period and we re-estimate the models based on the updated data sets. Finally we have 70 sets of one to weeks ahead forecast. We generate forecasts using the actual value of the explanatory variables and the forecasted values of the lagged dependent variables when the lead times are greater than one. The promotional variables are usually known to the retailer as they form part of an agreed promotional plan with suppliers. We specify the ADL models with the data from week 1 to week 200, which represents the model that would ideally be selected based on a foreknowledge of the data ([Fildes et al., 2011](#_ENREF_30)). Alternatively, the models can be re-specified for each rolling event based on each the moving estimation window.

We evaluate the forecasting performance of the various models using five error measures: the MAE, the Mean Absolute Scaled Error (MASE), the MAPE, the symmetric Mean Absolute Percentage Error (sMAPE), and the Average Relative Mean Absolute Error (AvgRelMAE). The MAE has been widely used in practice but has been criticized for its limitation of being scale dependent, in which case the results for the product category with large sales volumes would dominate the overall results ([Chatfield, 1988](#_ENREF_13)). In this study, the MAE for data series calculated with forecast horizon for the rolling event is:

where is the actual value in the forecast period for data series based on the rolling event, and is the forecast value for data series based on the rolling event[[4]](#footnote-4).

The MASE was proposed by [Hyndman and Koehler (2006](#_ENREF_44)). It can be considered as a “weighted” mean of the MAE based on the variations of the sales data in the estimation period. The MASE calculated across data series with forecast horizon for the rolling event is:

where in the equation of , the numerator, , is the MAE for data series calculated with forecast horizon for the rolling event. The denominator is the sum of one-step-ahead errors by a no change naïve model in the estimation period. is the actual value of data series in the estimation period for the rolling event, and is the total number of observations in the estimation period. The MASE has good properties such as being robust to zero actual values and scale independent, but it puts more weights to the data series which are comparatively stable (e.g. given the same MAE, will be extremely large if the no change naïve model generates very small errors), which makes it vulnerable to outliers.

The MAPE is the error measure most widely used in practice ([Fildes and Goodwin, 2007](#_ENREF_27)). It penalizes the forecasts above actual values more heavily than the forecasts below actual values ([Armstrong and Collopy, 1992](#_ENREF_6)). The sMAPE was proposed to overcome this disadvantage ([Makridakis, 1993](#_ENREF_51)). The two error measures calculated for data series s with forecast horizon for the rolling event are shown as follows:

However, both percentage error measures including the MAPE and the sMAPE can be distorting when the actual values and the forecast values are relatively small compared to the forecast error, in which case the resulting percentage errors become extremely large ([Hyndman and Koehler, 2006](#_ENREF_44)). The sales at the UPC level exhibit high degree of variations due to seasonal effects, changing stages of product life cycle, and particularly promotional activities. Under such a circumstance, it is very likely to have large forecast errors associated with relatively low product sales, which makes the percentage based error measures less advisable in our context ([Davydenko and Fildes, 2013](#_ENREF_19)).

The four error measures are all approximations of the unknown loss function of the retailer, and they penalize the forecast errors from different perspectives. To make a fair comparison, we assess the overall forecasting performance of the candidate models by calculating the mean value of all the four error measures across rolling events and data series considering different forecasting horizons :

where , , , and are the error measures calculated across data series and rolling events based on forecast horizon (i.e. , , and =1, 4 and 12). We can test the statistical significance for the difference between the forecasting results of the various models using the Wilcoxon signed rank (SR) test. The Wilcoxon SR test can be considered as a non-parametric version of a paired sample *t*-test but does not assume the errors follow any specific distribution.

Considering the limitations of the four error measures, [Davydenko and Fildes (2013](#_ENREF_19)) recommended the AvgRelMAE, which is a geometric mean of the ratio of the MAE between the candidate model and the benchmark model. In this study, we take an average of the AvgRelMAE across all the rolling events (i.e. ) and data series with respect to forecast horizon :

where is the MAE of the candidate model for data series calculated with forecast horizon for the rolling event and is the MAE of the benchmark model for data series calculated with forecast horizon for the rolling event. is the AvgRelMAE calculated across data series and rolling events with respect to forecast horizon (i.e. , , and =1, 4 and 12). The AvgRelMAE has the advantages of being scale independent and robust to outliers, also with more straightforward interpretation: a value smaller than one indicates an improvement by the candidate model.

8 results

We investigate the models' relative forecasting performance under conditions of two dimensions which may influence the out­ come: (1) different forecast horizons; and (2) whether the focal product is being promoted. Earlier research by Gi.ir Ali et al. (2009) compared the forecasting performance of different methods for the promoted forecast periods and non-promoted forecast peri­ ods separately. Their regression tree model beat the base-times-lift benchmark model when the focal product is being promoted but is outperformed by the benchmark model when the focal product is not on promotion. Their explanation is that the sales of the focal product are relatively stable when the focal product is not on pro­ motion , and this stability would benefit simple models such as the exponential smoothing method . This explanation neglects the fact that, even during the periods when the focal product is not being promoted , its sales could also be driven by promotions of other competitive products. In this study, we therefore divide the fore­ cast period into promoted periods and non-promoted periods .

Table 3 exhibits the forecasting accuracy of the various models averaged over horizons from one to twelve weeks based on the var­ ious absolute error measures as well as the rank for each model. We conduct the Wilcoxon sign rank test to examine the significance of the difference between the models' forecasting performances .11 Thus we rank the models equally if their forecasting performances are not significantly different from each other. For the whole forecast period, the base-times-lift approach has better performance com­ pared to the SES method . These two benchmark models are both sig­ nificantly outperformed by the ADL-own model for all the error measures, which suggests that the ADL model captures the effects of the price and the promotional activities more effectively than the base-times-lift approach. The ADL model and the ADL-DI model both incorporate the competitive information and they significantly outperform the ADL-own model for all the error measures.

Table 3 also shows the forecasting performance of the various models for the promoted forecast period. The SES method and the base-times-lift approach are significantly outperformed by the ADL-own model, which is consistent with the result for the whole forecast period. However, the ADL model no longer outperforms the ADL-own model significantly when ranked by the MASE and

11 We adopt a 5% significance level.

the sMAPE. There are two possible reasons. First , the impact of the promotional activities of the competitive products is substantially weaker than the impact of the promotion activities of the focal prod­ uct, and may become submerged in the latter ( Hochet al.,1995 ).Sec­ ond, retailers benefit from the sales of the whole product category rather than individual brands or UPCs, and they tend to avoid simul­ taneously promoting a product with its main competitors, because this will not necessarily increase store sales (e.g. a large proportion of the sales increase come from brand switching) but definitely low­ er the profit margin (Gupt a, 1988; Van Heerde et al., 2003). As a re­ sult, when the focal product is being promoted, there tends to be limited variations in the missing competitive explanatory variables, which makes the ADL model generate similar forecasts with the ADL-own model. However, the ADL-DI model significantly outper­ forms the ADL-own model for all error measures even for the pro­ moted period. One explanation is that the diffusion indexes used in the ADL-DI model incorporates competitive information not only from the most relevant competitive explanatory variables but also by pooling across all the competitive explanatory variables.

For the non-promoted forecast period, Table 3 shows that the SES method has the poorest forecasting result, but the base­ times-lift approach has very good forecasting performance- it significantly outperforms the ADL-own model for all the error measures expect for the MASE. This is consistent with the findings by Gi.ir Ali et al. (2009) that when the focal product is not on pro­ motion, the base-times-lift approach can be hard to beat. Essen­ tially it uses only the data from the non-promoted periods to calculate the smoothing forecast, removing the promotional peaks. The ADL model outperforms the base-times-lift approach for the MASE but has comparable performance for all the other error mea­ sures. However, the ADL-DI model still significantly outperforms the base-times-lift approach for all the error measures.

Tables 4 and 5 show the forecasting performance of the various models for different forecast horizons and the corresponding ranks according to the Wilcoxon sign rank test. The results are consistent with the results we observe for the one to twelve-weeks-ahead forecast horizon.

Table 6 shows the AvgRelMAE of various candidate models for different forecast horizons. When we compare the candidate mod­ els to the benchmark base-times-lift approach, the values are all smaller than 1, which indicates that the ADL model, the ADL-DI model, and the ADL-own model all outperform the benchmark base-times-lift model. In addition, the improvements by these models become more substantial as the forecast horizon increases. For example, the AvgRelMAE for the ADL-DI model decreases from 0.861 to 0.746 as the forecast horizon increases from one week to one-to-twelve weeks. Table 6 also calculates the AvgRelMAE of the candidate models compared to the ADL-own model. The values for the ADL-DI model are all smaller than 1, which indicates that the ADL-DI model outperforms the ADL-own model for all forecast horizons. Again we see the improvements become more substan­ tial as the forecast horizon increases (from 0.953 to 0.931). The val­ ues for the ADL model are all smaller than 1 except for the one week forecast horizon (i.e. the value is 1.005). However, as the forecast horizon in creases, the value of the AvgRelMAE for the ADL model decreases below 1 (e .g. 0.982 and 0.957), which sug­ gests that it has superior forecasting performance than the ADL­ own model which just relies on the price and promotional infor­ mation of the focal product.

9 conclusion and future research

Today one of the main concerns of grocery retailers is to reduce stock-outs while controlling the safety stock level. Stock-outs lead to profit loss and dissatisfied customers, while over-stock incurs additional inventory costs. One of the solutions to overcome this tension relies on more accurate forecasts. In practice, many retail­ ers use the base-times-lift approach to forecast product sales at the UPC level. The approach is based on a simple method and takes into account the effect of promotions in an *ad hoc* way. Other researchers have proposed sophisticated data mining models and machine learning algorithm s, trying to capture the effect of promo­ tions more effectively (Cooper et al., 1999; Gi.ir Ali et al., 2009). However, these methods have several limitations. For example, they ignore the carryover effect of promotions and/or overlook the effect of competitive information. These models are also com­ plex and difficult to inter pret. They rely on expertise that may well not be available and the company instead substitutes judgment for more formal modelling efforts ( Fildes & Goodwin, 2007).

In this paper, we investigate the value of the promotional infor­ mation including competitive price and competitive promotions in forecasting retailer product sales at the UPC level. We propose a forecasting approach which consists of two main stages. In the first stage, we deal with the problem of too many explanatory variable associated with the retail data at the UPC level using two distinct methodologies . First we identify the most relevant competitive explanatory variables with variable selection methods. Alterna­ tively we pool information across all the competitive variables and condense them into a handful number of diffusion indexes at the cost of some information loss, based on factor analysis. In the second stage, we incorporate the identified most relevant compet­ itive explanatory variables and the constructed diffusion indexes into the Autoregressive Distributed Lag (ADL) model following a general-to-specific modelling strategy. The general-to-specific ADL model captures the carryover effect of marketing activi ties, and is easily operational in terms of the interpretation of empirical results. Managers can make inference about how the sales of the focal product are driven by marketing activities of the focal prod­ uct and other competitive products.

The ADL model and the AOL- DI model we propose in this study significantly outperform the two basic benchmark models and the ADL-own model which is constructed exclusively with the price and promotion information of the focal product. The improve­ ments in forecasting accuracy become more substantial as the forecast horizon increases. This result proves the value of using competitive information in forecasting retailer product sales at the UPC level. We have also investigated the forecasting perfor­ mance of the models considering whether or not the focal product is being promoted. In both cases our methods outperform the benchmark models.

There remains the potential to improve upon the approach that we present in this paper. First perhaps future research must ad­ dress is to identify the competitive products more effectively or taking into account information from experts perhaps within the organization . For example, we have included most products within each product category when implementing the variable selection methods and the factor analysis, and the uncertainty could poten­ tially be reduced if a "short list" of the main competitors for each item can be produced based on the market knowledge of category managers (Dekimp e & Hanssens, 2000). Also in this study we do not take into account the effect of advertisin g. Thus one possible way to improve the forecasting accuracy is to incorporate advertis­ ing information, although previous studies found the effect of advertising temporary and fragile (Chandy, Tellis, Macinnis, & Thaivanich , 2001). Datasets such as this may also contain evidence on the different types of promotions such as simple price reduction, bonus buy, and coupon, and it may be possible to distinguish between them. However, this would substantially increase the number of competitive explanatory variables, which adds more uncertainty to the selection of the competitive explanatory vari­ ables and the construction of the diffusion indexes. An additional possible way to improve the model is to further incorporate in format ion from other substitutive and complementary product categories ( Bandyopadhyay, 2009; Kamakura & Kang, 2007; Song & Chintagunta, 2006). Again the effect would be to increase the size of the variable set dramatically and demand even more of the practitioner.

Two other modelling issues might merit further work, an exam­ ination of the exogeneity of the promotional variables (though this would be unlikely to lead to forecast improvements) and the use of Bayesian estimation methods. This study was carried out using weekly aggregate data (across stores). While such forecasts are needed for ordering and distribution decision making, store-level forecasts on a daily basis are also needed. They can of course be de­ rived from simple proportionate disaggregation processes, an ap­ proach often used in practice, but it is an open research question as to whether daily explicit model-based disaggregate forecasts are to be preferred.

In this study, we constructed the general-to-specific AOL model manua lly. The modelling process is subjective and relies overmuch

on our tacit knowledge of modelling specification. Alternatively, Hendry and Krolzig (2001) proposed the PcGive software which automatically constructs the model. The software starts with a general model and simplifies the model exclusively based on diag­ nostic tests. However, the software does not incorporate marketing theory and Fildes et al. (2011) found that the ADL model built man­ ually by the model builder outperformed the one constructed by the PcGive software.

Lastly, our methods, as well as all the other forecasting methods in previous studies, are built on the assumption of a "stationary" market where all the other influencing factors including economic conditions, consumer preferences, future advertising campaigns, new product entry, etc. are assumed to be constant (Ehrenberg , 1988). However, this assumption cannot strictly hold in reality. As a result, the effectiveness of the included explanatory variables may change over time (take price as an example, customers may become more price/promotion sensitive during an economic crunch ). Thus we may potentially improve forecasting accuracy by either taking into account the missing market information (e.g. economic factor , manufacturers' future plans for advertising campaigns, and customer preferences, etc.) or trying to capture how the effectiveness of the included explanatory variables change over tim e, which is the subject of further research. In practical terms however most of these missing variables are slowly chang­ ing and the practical expedient of taking a rolling sample updating the models should work well.

The industry standard benchmark method proved of lim ited va­ lue for capturing promotional effects. This paper shows that for practical purposes the company forecaster could achieve superior forecasting performance by incorporating competitive in formation , either through variable selection methods or factor analysis, with the ADL model built manually following the general-to-specific strategy. Although the development of an automatic ADL model proved costly in terms of forecasting accuracy, and we see this as difficult to implement for most organizations where there might not be experts in econometric model building ( Montgomery, 2005), the benefits of embracing a more sophisticated model build­ ing approach have proved substantial. For example, the ADL-DI model has reduced the sMAPE by around 30% compared to the bases-times-lift model for the one-to-twelve weeks forecast hori­ zon and if implemented should lead to substantial savings in the distribut ion chain. At Whether collaborative forecasting, as part of a Sales and Operations Planning Process (S&OP), could further enhance the gains from the statistical models proposed here re­ mains an open research question.

Acknowledgement

Thanks are given to two anonymous reviewers for their valu­ able comments which improve the earlier version of this manu­ script. In add ition, the authors wish to thank Kilts Center for Marketing (The University of Chicago Booth School of Business) for making available the data.

Reference

1. The calendar events include *Halloween*, *Thanksgiving*, *Christmas*, *New Year’s Day*, *President’s Day*, *Easter*, *Memorial Day*, *4th of July*, and *Labour Day*. [↑](#footnote-ref-1)
2. In the preliminary analysis, *L* is initially set as two. If the general model does not pass the misspecification tests, more lags of the price, promotion, and sales variables are added to the general model. In our modelling, for most UPCs, the ADL models do not contain more than two lags of these variables. [↑](#footnote-ref-2)
3. We choose to retain four diffusion indexes for competitive prices and four diffusion indexes for competitive promotions based on preliminary analysis. For each product category, the percentages of explained variation in the competitive price data series range from 51% to 79%, and the percentages of explained variation in the competitive promotion data series range from 32% to 69%. [↑](#footnote-ref-3)
4. Note that, in this study, although our econometric models are based on log sales, we calculate all the error measures after we transform them back to original levels. [↑](#footnote-ref-4)