

Manuscript Number: EJOR-D-18-00185

Title: Forecasting Retailer Product Sales in The Presence of Structural Breaks

Article Type: Innovative Application of OR

Section/Category: (T) Forecasting

Keywords: Forecasting; OR in marketing; Analytics; Retailing

Corresponding Author: Dr. tao huang, Ph.D

Corresponding Author's Institution: University of Surrey

First Author: tao huang, Ph.D

Order of Authors: tao huang, Ph.D; Robert Fildes, PhD; Didier Soopramanien, PhD

Abstract: Grocery retailers need accurate forecasts at SKU level for their inventory management decisions. Previous studies have developed forecasting models which incorporate the effect of various marketing activities including prices and promotions. These models, however, do not consider that the effect of these marketing activities on product sales may not be constant over time. Under such a circumstance, the models could be subject to the structural break problem, i.e., the models with constant parameters are unable to capture the varying effect of the marketing activities. As a result, the generated forecasts may potentially be biased and less accurate. In this study, we propose new forecasting methods for retail product sales by taking into account the problem of structural break. Our methods generate more accurate forecasts compared to conventional models which assume constant parameters for various marketing activities.

Dear Sir or Madam,

We have corrected the following according to the comment in the Technical Check Results:

HIGHLIGHTS

Please ensure there are no acronyms, abbreviations or formulae in your highlights whilst not exceeding the limit of 85 characters per bullet point (including spaces).

We have corrected the highlights and we ensure that there are no acronyms, abbreviations, and the highlights now do not exceed the limit of 85 characters per bullet point (including spaces).

Abstract

Define the acronym SKU on its first time of use, and then again in the abstract.

We have defined the acronym SKU on its first time of use and then again the abstract.

Thank you!

Best regards,

Tao Huang

Retailer product sales from a wide range of product categories

Forecasting retailer product sales by taking into account structural breaks

Explore the determinants of improved forecasting performance

Forecasting Retailer Product Sales in The Presence of Structural Breaks

Tao Huang¹

Surrey Business School, University of Surrey, GU2 7XH, UK

Robert Fildes

Centre for Marketing Analytics and Forecasting, Lancaster University, LA1 4YX, UK

Didier Soopramanien

Lancaster University management school, Lancaster University, UK, LA1 4YX

Abstract

Grocery retailers need accurate forecasts at Stock Keeping Unit level for their inventory management decisions. Previous studies have developed forecasting models which incorporate the effect of various marketing activities including prices and promotions. These models, however, do not consider that the effect of these marketing activities on product sales may not be constant over time. Under such a circumstance, the models could be subject to the structural break problem, i.e., the models with constant parameters are unable to capture the varying effect of the marketing activities. As a result, the generated forecasts may potentially be biased and less accurate. In this study, we propose new forecasting methods for retail product sales by taking into account the problem of structural breaks. Our methods generate more accurate forecasts compared to conventional models which assume constant parameters for various marketing activities.

Keywords:

Forecasting, OR in marketing, Analytics, Retailing

¹Corresponding author at Surrey Business School, University of Surrey, GU2 7XH, UK. Tel: 01483 68 6359, email: t.huang@surrey.ac.uk; r.fildes@lancaster.ac.uk (r.Fildes); d.Soopramanien@lancaster.ac.uk (d.soopramanein)

1. Introduction

Grocery retailers rely on accurate sales forecasts for their inventory management (Petropoulos, Makridakis, Assimakopoulos, & Nikolopoulos, 2014). Poor forecasts of product sales lead to poor service arising from out-of-stock conditions or, alternatively, inflated costs due to overstocking. When a specific item is out-of-stock, retailers directly lose the income and profit from the sale of the item. If the out of stock situation happens on a regular basis, it can lead to consumer dissatisfaction. In the long term, retailers may see customers switching to other retail chains (Corsten & Gruen, 2003). To avoid such situations, retailers may intentionally overstock to maintain a high customer satisfaction level but this significantly raises inventory costs (e.g., capital cost, warehousing, and deterioration etc.) and reduces profits (Cooper, Baron, Levy, Swisher, & Gogos, 1999). In 2014, retailers in North America had a loss of \$634.1 billion due to out-of-stock and spent \$471.9 billion on overstock (OrderDynamics, 2015). One of the solutions to mitigate the dilemma is to generate more accurate sales forecasts at Stock Keeping Unit (SKU) level, which improves the effectiveness of the supply chain management by reducing the bullwhip effect and enabling the Just-In-Time delivery (Ouyang, 2007; Sodhi & Tang, 2011)².

In practice, many retailers generate forecasts at SKU level using a two-stage ‘base-lift’ approach. The forecasts are generated separately depending on whether or not the focal product is being promoted. The ‘base’ forecasts are generated using simple univariate models, while the ‘lift’ effect, which is effectively caused by marketing activities including price reductions and promotions, is estimated by the brand/category manager based on his/her experience. In this context, some previous studies have proposed procedures to help managers improve the accuracy of their judgments (e.g., Fildes, Nikolopoulos, Crone, & Syntetos, 2008; Goodwin, 2002; Nikolopoulos, 2010). Others have developed models to estimate the ‘lift’ effect based on data (Cooper et al., 1999; Cooper & Giuffrida, 2000; Trusov, Bodapati, & Cooper, 2006). The third type of approach develops methods to directly generate the final forecast of the product sales. For example, Gür Ali, SayIn, van Woensel, and Fransoo (2009) proposed the regression tree method with a range of variables constructed from the sales, price, and promotion of the focal product. Huang, Fildes, and Soopramanien (2014) proposed two-stage general-to-specific Autoregressive Distributed Lag (ADL) models which incorporated the promotional information of not only the focal product but also of the competitive products within the same product category. Ma, Fildes, and Huang (2016) further integrated the promotional information of the products from related product categories.

² A stock keeping unit (SKU) is a product item which can be individually identified by the business’ inventory system (Sawaya & Giauque, 1986)

1 However, all these studies assume that the impact of marketing activities on product sales remains
2 constant over time. In practice, the effect of prices and promotions may change due to the many non-
3 controllable factors which may include, for instance, changing economic conditions, changes in
4 consumer tastes, and the entry of new competitors etc., some of which are neither observable or
5 measurable (Wildt, 1976; Wildt & Winer, 1983). Customers may become more sensitive to prices and
6 promotions during an economic crunch. They may change their tastes due to factors including their
7 familiarity with the product, and their changing lifestyle and social status (Meeran, Jahanbin,
8 Goodwin, & Quariguasi Frota Neto, 2017). When a new competitor enters the market, the effect of
9 prices and promotions of the focal product may be reduced not only because the new competitor
10 launches their marketing activities but also because customers seek variety. In the year of 2014, the
11 German low-price retail chain Aldi opened more than 400 stores in the United States, leading to
12 changes in customer grocery purchasing habits, which then put pressures on existing retail chains
13 (Loeb, 2015).

21 Under any of the circumstances described above, conventional models which use constant parameters
22 to represent the effect of the price and promotions may potentially be subject to the structural break
23 problem (Allen & Fildes, 2001; Armstrong, 2001). The model which is subject to structural break
24 may generate biased and less accurate forecasts. The structural break problem has been historically
25 addressed in the macroeconomics literature (see Clements & Hendry, 1994; Pesaran & Timmermann,
26 2005). As an example, Ang and Bekaert (2002) explored the change of the effect of the financial
27 interest rate on stock market returns due to exogenous factors including in market sentiment shifts and
28 new regulations. The problem of the structural break has been totally overlooked in forecasting
29 retailer product sales. In this study, we examine new effective methods to forecast retailer product
30 sales by taking into account the structural break problem.

41 Our research in the domain of retail forecasting in particular at SKU level is significant for the
42 following reasons. First, our research is the first research which investigates the structural break
43 problem in forecasting retailer product sales. The data in retailer product sales at SKU level exhibit
44 unique characteristics compared to data in other areas (e.g., macroeconomics). Also, the methods
45 which deal with the structural break problem by reducing the associated forecast bias come with the
46 cost of inflated forecast error variance (which also affects the forecasting accuracy, as discussed in
47 later sections). Under such circumstances, whether or not we can improve the forecasting accuracy by
48 dealing with the structural break problem becomes an empirical question. The final results indicate
49 that our models have superior forecasting performance compared to conventional models which
50 assume no change in the effect of product prices and promotions. Second, unlike any earlier studies
51 which rely on incorporating additional information on the marketing mix (which leads to additional
52 cost), our methods rely on how limited promotional information could be effectively utilized. In

practice, the change of the effect of the marketing activities may be caused by many factors (as mention above) for which the data are difficult or infeasible to collect or measure. Therefore, our methods add values without incurring additional costs to retailers. Third, our research provides an evaluation of various forecasting methods which offers operational guidance to not only retailers but also to manufacturers when competitive promotional information is unavailable. Fourth, the methods we propose are fully automatic and easy to implement compared to Huang et al. (2014).

The remainder of the paper is organized as follows: section 2 summarizes previous studies in the literature related to forecasting retailer product sales and the change of the effect of marketing activities. Section 3 explains the origins and the consequence of the structural break problem. In section 4, we introduce two methods which are used in the macroeconomics area to deal with the structural break problem. Section 5 explores the data. In section 6, we propose our new three-stage forecasting methods. Section 7 describes the design of the model evaluation. Section 8 summarizes and discusses the evaluation results in order to provide a convincing demonstration of their performance. In Section 9, we explore the characteristics of the situations where the proposed models garner the greatest improvements. In the last section, we make recommendations for both manufacturers and retailers, address research limitations, and highlight directions for future research.

2. Literature review

In practice, many retailers forecast their product sales at SKU level using a two-stage ‘base-lift’ method. The method entails dividing the data into promoted and non-promoted periods based on whether the focal SKU is being promoted. The method is a combination of simple univariate methods (for the non-promoted period) and human judgments by brand/category managers (for the promoted period) (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Fildes et al., 2008). A number of studies have been devoted to helping managers with better adjustment procedures by overcoming their cognitive biases (Lee, Goodwin, Fildes, Nikolopoulos, & Lawrence, 2007; Petropoulos, Fildes, & Goodwin, 2016). Other studies try to improve the adjustment with model-based forecasting systems. e.g., they estimate the ‘lift’ effect by the promotional event based on information related to previous promotions, store/category features, and manufacturers etc. (Cooper et al., 1999; Cooper & Giuffrida, 2000; Trusov et al., 2006). One limitation of these methods is that, as they split the data into two periods, they tend to overlook the information in the promoted period when forecasting the product sales in the non-promoted period, and vice versa.

Other studies have proposed more holistic methods to generate the forecasts. Divakar et al. (2005) developed the CHAN4CAST system with models of dynamic regression structures to forecast brand volume sales for the manufacturer/channel. Gür Ali et al. (2009) evaluated the forecasting

performance of support vector machine (SVM) models and regression tree models. Huang et al. (2014) proposed two-stage general-to-specific ADL models which incorporate competitive promotional information within the same product category of the focal product. Ma et al. (2016) further integrated the promotional information not only from the same category but also from other related categories. These studies tried to generate accurate sales forecast by capturing the effect of marketing activities. For example, the short-term effect of prices and promotions (Blattberg, Briesch, & Fox, 1995), the (asymmetrical) competitive effect (Andrews, Currim, Leeflang, & Lim, 2008; Dekimpe, Hanssens, & Silva-Risso, 1999; Wedel & Zhang, 2004; Wittink, Addona, Hawkes, & Porter, 1988), and the dynamic effects which lead to purchase acceleration and anticipation (Mace & Neslin, 2004; Van Heerde, Gupta, & Wittink, 2003).

The studies above all assume the constant effect of marketing activities. However, evidence has accumulated and shows that the effect of marketing activities including prices and promotions may change over time (e.g. Houston & Weiss, 1975; Little, 1966; Mahajan, Bretschneider, & Bradford, 1980; Moïn pour, McCullough, & MacLachlan, 1976; Monroe & Guiltinan, 1975; Morrison, 1966; Myers, 1971; Myers & Nicosia, 1970; Wichern & Jones, 1977; Wildt, 1976; Winer, 1979). Wildt (1976) and Wildt and Winer (1983) attribute the change in the effect of the marketing activities to the change in economic conditions, consumer tastes, and the competitive environment etc. Customers may find price reductions and promotions more attractive when there is an economic crunch compared to other time periods. Customers may also display a change in their tastes and preferences. This can occur when customers accumulate more knowledge of the product, when they seek variety, and when they reach a different social status and decide to adopt a different lifestyle (Meeran et al., 2017). These individual changes may then lead to substantial aggregate effects. Research at store level finds that the introduction of new brands in a product category (e.g., the store-owned brand) decreases the promotional elasticities of premium national brands and increase promotional elasticities of the second tier national brands (Nijs, Dekimpe, Steenkamps, & Hanssens, 2001; Van Heerde, Srinivasan, & Dekimpe, 2008). Lastly, the effect of prices and promotions may change during the different stages of the product lifecycle (Mahajan et al., 1980). Overall, changes in the effect of marketing activities and consumer responses on sales, however, has been overlooked by previous studies in the forecasting literature.

3. Structural break and potential forecast bias

Conventional models with constant parameters tend to overlook the change in the effect of the marketing activities at different points in time such as price changes and promotions on product sales. As a result, the generated forecasts will potentially be biased and less accurate (Allen & Fildes, 2001; Armstrong, 2001). This is referred to as the structural break problem and has been addressed in the

macroeconomics literature (Castle, Doornik, & Hendry, 2008; Hendry, 2018; Pesaran & Timmermann, 2007). Pesaran and Timmermann (2005) demonstrate analytically how a structural break may lead to forecast bias using a simple regression model.

In a retailer context, suppose that we have the sales and price information of the product from week 1 to week T , i.e., $[1:T]$, and, for exposition, we presume that the price is the only factor available to us and there is a structural break at week T_1 (where $1 < T_1 < T$). This structure break may be caused by other factors such as economic crunch, change of consumer taste, or new competitor entry etc as introduced in the previous section. Thus, the true parameter of the price variable changes from β_1 to β_2 after T_1 . The unobserved true demand can be represented as follows:

$$y_t = 1_{\{t \leq T_1\}} \beta_1' x_t + (1 - 1_{\{t \leq T_1\}}) \beta_2' x_t + u_t$$

where, $1_{\{t \leq T_1\}}$ is an indicator which equals to 1 before week T_1 and 0 afterwards. y_t and x_t are the sales and the price of the product at week t . We consider x_t to be strictly exogenous as we assume that retailers do not change product prices based on their short-term sales³. u_t is the error term, and we assume that $u_t \sim iid(0, \sigma_1^2)$ when $t \leq T_1$ and $u_t \sim iid(0, \sigma_2^2)$ when $t > T_1$. We may estimate a model with a functional form which is congruent with the demand (e.g., $\hat{y}_t = \hat{\beta} x_t + u_t$) using the data before and after the structural break, e.g., $[m:T]$, where $1 \leq m < T_1 < T$. The OLS estimate for the parameter is:

$$\hat{\beta}_T(m) = (X'_{m,T} X_{m,T})^{-1} X'_{m,T} Y_{m,T}$$

where $X_{m,T}$ and $Y_{m,T}$ are the matrices of the sales and price variable for the time period from week m to week T . We assume that there is no structural break after week T and the true demand after week T remains as $y_t = \beta_2' x_t + u_t$, when $t > T$. Therefore, the h -step ahead forecast error at week $T+h$ can be represented as:

$$\begin{aligned} \hat{e}_{T+h}(m) &= (\beta_2 - \hat{\beta}_T(m))' x_{T+h} + u_{T+h} \\ &= (\beta_2 - (X'_{m,T} X_{m,T})^{-1} X'_{m,T} Y_{m,T})' x_{T+h} + u_{T+h} \\ &= (\beta_2 - \beta_1)' X'_{m,T_1} X_{m,T_1} (X'_{m,T_1} X_{m,T_1})^{-1} X_{T+h} - u'_{m,T} X_{m,T} (X'_{m,T_1} X_{m,T_1})^{-1} X_{T+h} + u_{T+h} \\ &= (\beta_2 - \beta_1)' Q_{m,T_1} Q_{m,T_1}^{-1} X_{T+h} - u'_{m,T} X_{m,T} Q_{m,T_1}^{-1} X_{T+h} + u_{T+h} \end{aligned}$$

where $Q_{m,T_1} = X'_{m,T_1} X_{m,T_1}$, and u_{T+h} is the matrix for the error term at week $T+h$.

³ Analytical evidence for the models with endogenous explanatory variables can be found in Clements and Hendry (1999) and Pesaran and Timmerman (2005, 2007).

Therefore, the forecast at week $T + h$ is biased as $E[\hat{e}_{T+h}(m)|x_{T+h}] = (\beta_2 - \beta_1)'Q_{m,T_1}Q_{m,T_1}^{-1}X_{T+h}$, and it is unequal to zero⁴.

4. Dealing with structural breaks

The bias due to the structural break may be mitigated by specifying non-zero values for the model's errors in the forecasting period, which is referred as the Intercept Correction (IC) method. (Clark & McCracken, 2007; Clements & Hendry, 1994; Clements & Hendry, 1999). For example, if we believe that the model is subject to structural break and forecasts are biased, we may estimate the bias as the average value of the most recent residuals, i.e., $\widehat{Bias}_{IC} = \sum_{\omega}^W \hat{e}_{T-\omega} = \sum_{\omega}^W [(\beta_2 - \hat{\beta}_T(m))' x_{T-\omega} + u_{T-\omega}]$, where W is the number of residuals being used to estimate the forecast bias. When $W = 1$, the estimate reduces to \hat{e}_{T-1} , which is the residual at the forecast origin (e.g., Chevillon, 2016). The estimated bias are then added back to the out-of-sample forecasts, which may potentially improve the forecasting accuracy but at a cost of inflated forecasting error variance (Clements & Hendry, 1999). In a retailer context, product sales at SKU level often exhibit large variations, unexpected outliers, and missing values, which makes estimating the forecast bias a difficult task.

An alternative method is to combine the forecasts which are generated by the same model but with different estimation windows while expecting a trade-off between reduced forecast bias and potentially increased forecast error variance (Pesaran & Timmermann, 2005; Pesaran & Pick, 2011). Ideally, if we know that there exists a structural break and it occurs at T_1 , we may estimate the model exclusively with the post-break data, i.e., $[T_1 : T]$, and generate unbiased forecasts. However, as we do not know the location of the break, we may estimate the model using the data which are closest to the forecast origin (e.g., we keep m as large as possible) in conformity with maintaining the degrees of freedom so that there are enough observations to estimate the model. If m by chance becomes larger than T_1 , the model will be estimated with the post-break data only and will generate unbiased forecasts. However, this does not necessarily lead to improved forecasting accuracy because the forecasting error variance would increase due to smaller estimation window (i.e., we are using fewer observations to estimate the model). The Mean Squared Error (MSE) at week $T + h$ can be represented as $MSE(T + h, m|X_{T+h}) = E(\hat{e}_{T+h}^2(m)|X_{T+h}) = \sigma_2^2(1 + B_m + E_m)$, where $B_m = \mu'Q_{m,T_1}Q_{m,T}^{-1}x_{T+h}x_{T+h}'Q_{m,T}^{-1}Q_{m,T_1}\mu$, and can be interpreted as the squared forecast bias; $E_m = x_{T+h}'Q_{m,T}^{-1}x_{T+h} + \psi(x_{T+h}'Q_{m,T}^{-1}Q_{m,T_1}Q_{m,T_1}^{-1}x_{T+h})$, and can be interpreted as the efficiency term ($\sigma_2^2E_m$ is the forecasting error variance), $\mu = (\beta_2 - \beta_1)/\sigma_2$, $Q_{m,T} = X_{m,T}'X_{m,T}$, and $\psi = (\sigma_1^2 - \sigma_2^2)/\sigma_2^2$.

⁴ In Appendix A in the supplementary material, we illustrate the impact of the structural break on the forecasting performance using a simulation example.

Pesaran and Timmermann (2005) show analytically that the change of the *MSE* for week $T + h$ when we estimate the model with data $[m + 1, T]$ compared to with the data $[m, T]$ is:

$$\begin{aligned}\Delta_{m+1,m} &= MSE(T + h, m + 1|X_T) - MSE(T + h, m|X_T) \\ &= \sigma_2^2[(B_{m+1} - B_m) + (E_{m+1} - E_m)]\end{aligned}$$

where $MSE(T + h, m + 1|X_T)$ is the *MSE* at week $T + h$ based on the estimation window $[m+1, T]$. When the observation at week m is excluded in the estimation, the change of the squared bias term (e.g., $B_{m+1} - B_m$) will always be non-positive (i.e., the bias will decrease), but the change of the efficiency term (e.g., $E_{m+1} - E_m$) depends on the error variance before and after the structural break. If $\sigma_1^2 \geq \sigma_2^2$ (e.g., there are more pre-break variations compared to post-break variations in the product sales which cannot be explained by the price variable), E_{m+1} will be smaller than or equal to E_m , and the *MSE* will decrease as the change for both the squared bias term and the efficiency term are non-positive. However, if $\sigma_1^2 \leq \sigma_2^2$, E_{m+1} will be larger or equal to E_m . Under this condition, the *MSE* may either increase or decrease depending on how the non-positive change of the squared bias term compares to the non-negative change of the efficiency term. As a result, when we exclude pre-break data and adopt a smaller estimation window, we may have either better or worse forecasting performance depending on the trade-off between the reduced forecast bias and the potentially inflated forecasting error variance. Therefore, the forecasts generated by the model with larger estimation windows may be subject to larger bias (contains more pre-break data) but associated with smaller forecast error variance (with more observations), and vice versa. As it is difficult to find the location of the structural break, we can combine the forecasts generated by the models with different estimation windows, which may potentially lead to higher forecasting accuracy by making an effective trade-off between the forecast bias and the forecasting error variance (Clemen, 1989; Jose & Winkler, 2008).

For example, we may combine the forecasts with equal weights as it has been found effective and easy to implement.(Clements & Hendry, 1998; Dekker, van Donselaar, & Ouwehand, 2004; Fildes & Stekler, 2002; Pesaran, Schuermann, & Smith, 2009). We may estimate the model using the most recent ω observations to generate the 1st set of the h -step-ahead forecast, e.g., $\hat{y}_{T+h,1} = X'_{T+h}\hat{\beta}_{T-\omega:T}$, where $\hat{\beta}_{T-\omega:T}$ represents the parameters estimated with the sample window $[T - \omega, T]$. The value of ω is arbitrarily chosen given there are enough observations to estimate the model and there are enough variations in the explanatory variables. We then add more observations (e.g., one) to the estimation window and generate the 2nd set of the h -step-ahead forecast, e.g., $\hat{y}_{T+h,2} = X'_{T+h}\hat{\beta}_{T-\omega-1:T}$ and so forth. We may have the $(T - \omega + 1)^{th}$ set of the h -step-ahead forecasts, e.g., $\hat{y}_{T+h,T-\omega+1} =$

$X'_{T+h}\hat{\beta}_{1:T}$. We calculate the final forecasts as the average value of the $(T - \omega + 1)$ sets of h -step-ahead forecasts:

$$\hat{y}_{T+h}(T, \omega) = (T - \omega + 1)^{-1} \sum_{m=1}^{T-\omega+1} \hat{y}_{T+h,m} = (T - \omega + 1)^{-1} \sum_{m=1}^{T-\omega+1} X'_{T+h}\hat{\beta}_{m:T}$$

This method is referred as the Estimation Window Combining (EWC) method and has been used to help VAR models forecast financial variables (Pesaran et al., 2009)⁵.

5. The data

In this study, we evaluate the forecasting performance of various models using the retail dataset made available by the Information Resources, Inc. (IRI) company. A description of the dataset can be found in Bronnenberg, Kruger, and Mela (2008). The dataset contains weekly data at SKU level with variables including product unit sales, price, features, and displays etc. We conduct our evaluation based on 1831 SKU's for 28 product categories from 28 different stores. Table 1 shows the basic statistics of the selected SKU's during a period of 202 weeks for each product category⁶. Some product categories (e.g., Carbonated Beverages and Hotdog) exhibit much higher promotional intensity compared to others (e.g., Margarine/Butter and Mayonnaise). Figure 1 exhibits the data series for a typical SKU in the Beer category as an example: it indicates that sales spikes are usually associated with the price reductions and feature/display promotions of the focal product, as well as calendar events (e.g., Halloween, Thanksgiving, and Christmas etc.).

Table 1. Statistical description of different product categories

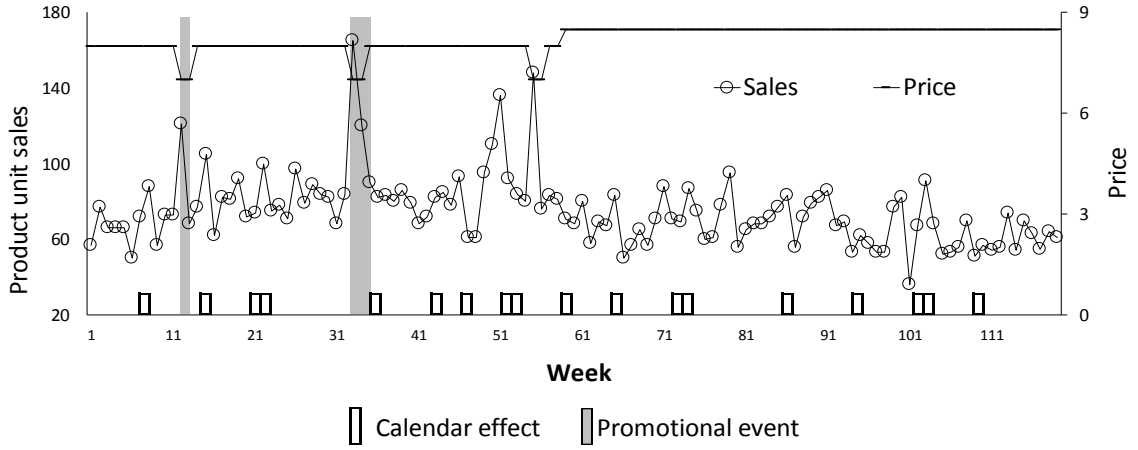
Category	Price mean	Sales mean	Display percentage	Feature percentage	Number of SKU's
Beer	8.3	20.6	13.90%	4.00%	169
Blades	8.1	14.6	7.40%	2.20%	22
Carbonated Beverages	2.1	113.6	26.80%	15.60%	82
Cigarette	22.3	22.2	0.00%	0.80%	202
Coffee	5.2	14.5	5.20%	2.90%	86
Cold cereal	3.5	70.7	4.00%	18.10%	125
Deodorant	2.7	6.9	4.10%	5.20%	126
Face Tissue	2.1	75.8	0.30%	11.70%	6
Frozen Dinner	2	43.8	5.30%	23.70%	87
Frozen pizza	3.4	31.2	8.90%	9.10%	147
Household Cleaner	2.5	29.9	0.30%	3.60%	18
Hotdog	4	68.6	13.20%	15.60%	35
Laundry Detergent	8.8	28.9	2.30%	8.80%	57
Margarine/Butter	2	71.4	0.10%	6.30%	36
Mayonnaise	3	79.7	3.00%	0.40%	22
Milk	2.5	222.3	2.10%	1.80%	30

⁵ In Appendix B in the supplementary material, we demonstrate how we can achieve more accurate forecasts with the IC method and the EWC method using simulation examples.

⁶ We select the SKUs with positive movements for at least 90% of the time.

Mustard & Ketchup	2.1	64.5	5.30%	0.90%	22
Peanut butter	3.7	34.2	3.20%	0.60%	15
Photo	7.2	9.2	4.60%	5.10%	13
Salty snacks	2.3	50.9	6.70%	5.00%	101
Shampoo	3.5	9.9	12.80%	7.10%	70
Soup	1.5	61.6	1.20%	9.70%	139
Spaghetti sauce	2.4	39.1	1.60%	6.50%	52
Sugar substitutes	2.8	14.5	0.10%	1.40%	20
Toilet Tissue	5.4	89.1	4.30%	8.30%	20
Toothbrush	2.6	8.7	3.10%	6.30%	28
Toothpaste	2.8	35.5	11.00%	12.50%	25
Yogurt	1.1	115.1	0.70%	6.30%	75

Figure 1. Store level data for an SKU in the Beer category⁷



6. The models

In this study, we propose new forecasting methods with three stages. At the first stage, we identify the most informative competitive explanatory variables for the focal product. Grocery retailers typically sell hundreds of SKU's in a typical product category and this leads to hundreds of potential competitive explanatory variables for the focal product. Incorporating all the variables into the model would easily overfit the model and even make the estimation infeasible (Martin & Kolassa, 2009). Therefore, we initially select the most relevant variables using the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996). For example, we have the following model for the sales of a specific SKU:

$$\ln(y_{0,t}) = X\beta + u, \text{ subject to } \sum_{j=1}^N |\beta_j| = \eta, \eta \leq \eta_0$$

⁷ In Figure 6, the calendar events include Halloween, Thanksgiving, Christmas, New Year's Day, President's Day, Easter, Memorial Day, the 4th of July, and Labour Day. The promotional events include Feature and Display.

where $\ln(y_{0,t})$ represents log product sales of the focal product at week t .

X is the matrix for the explanatory variables including product prices, features, and displays of all the products in the same product category.

u represents the identically distributed error term.

β represents the vector for the parameter coefficients.

N is the total number of SKUs for the category.

η_0 is the shrinkage factor.

The LASSO procedure imposes a constraint to the sum of the absolute values of the models' parameter coefficients. It removes less relevant explanatory variables by pushing their parameter coefficients towards zero. We control the model simplification process using the shrinkage factor based on 10-fold cross validation (Ma & Fildes, 2017; Ma et al., 2016)⁸.

At the second stage, we construct the General Autoregressive Distributive Lag (ADL) model based on the variables retained by the LASSO procedure with their dynamic terms (Huang et al. 2014). One limitation of the LASSO procedure is that it may potentially miss important variables under the condition of high multicollinearity (Fan & Lv, 2008; Ma et al., 2016). In practice, retailers tend to promote relevant products at the same time, which may even increase the multicollinearity. Therefore, we include the marketing variables of the focal product in the general ADL model. The general ADL model takes into account the dynamic effect of the (LASSO retained) marketing activity variables as well as a term capturing the potential trend, four-week seasonality, and calendar events. The general ADL model can be represented as:

$$\begin{aligned} \ln(y_{0,t}) = & \text{intercept} + \tau * \text{time} + \sum_{j=1}^L \alpha_j \ln(y_{0,t-j}) + \sum_{j=0}^L \beta_{0,j} \ln(p_{0,t-j}) + \sum_{j=0}^L \gamma_{0,j} \text{Feature}_{0,t-j} \\ & + \sum_{j=0}^L \gamma_{0,j} \text{Display}_{0,t-j} + \sum_{m=1}^M \sum_{j=0}^L \beta_{m,j} \ln(p_{m,t-j}) \\ & + \sum_{n=1}^N \sum_{j=0}^L \gamma_{n,j} \text{Feature}_{n,t-j} + \sum_{n=1}^P \sum_{j=0}^L \gamma_{n,j} \text{Display}_{n,t-j} + \sum_{d=1}^{12} \theta_d \text{Four_week_dummy}_d \\ & + \sum_{c=1}^9 \sum_{v=0}^1 \delta_{c,v} \text{CalendarEvent}_{c,t-v} + \varepsilon_t \end{aligned}$$

where $\ln(y_{0,t})$ is the log sales of the focal product at week t .

time is the term which captures any potential trend during the estimation period (Song & Witt, 2003).

⁸ Alternative schemes including information criteria are also available (e.g., Huang et al., 2014). We find little difference between the results by these schemes.

$\ln(p_{0,t-j})$ and $\ln(p_{m,t-j})$ represent the log price of the focal product and a competitive product, m , at week $t - j$.

$Feature_{0,t-j}$ and $Display_{0,t-j}$ represents the Feature dummy for the focal product at week $t - j$.

$Four_week_dummy_d$ is the d^{th} four-week-dummy variable.

$CalendarEvent_{c,t-v}$ is the dummy variable for the c^{th} calendar event at week $t - v$. The dummy variable represents the week of the calendar 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 events⁹.

$\alpha_j, \beta_{0,j}, \gamma_{0,j}, \beta_{m,j}, \gamma_{m,j}, \theta_d, \delta_{c,v}, \tau$ are the parameters.

ε_t is the error term and we assume $\varepsilon_t \sim iid(0, \sigma^2)$.

L is the order of the lags and is set as 2.

M , N , and P are the numbers of selected competitive price, Feature, and Display variables for the product category.

We then simplify the general ADL model using the LASSO procedure (we refer to this simplified model as the ADL-raw model thereafter). Previous studies indicate that models simplified by the LASSO procedure have good forecasting performance and outperform traditional models specified based on statistical significance (Epprecht, Guegan, & Veiga, 2013; Ma et al., 2016). The LASSO procedure also enables the automation of the statistical forecasting task which becomes essential as typically grocery retailers stock a tremendous number of SKUs (Cooper et al., 1999). However, to mitigate the LASSO procedure's limitation of missing important variables due to multicollinearity, we construct the following supplementary parallel model which only includes the price and promotion variables of the focal product:

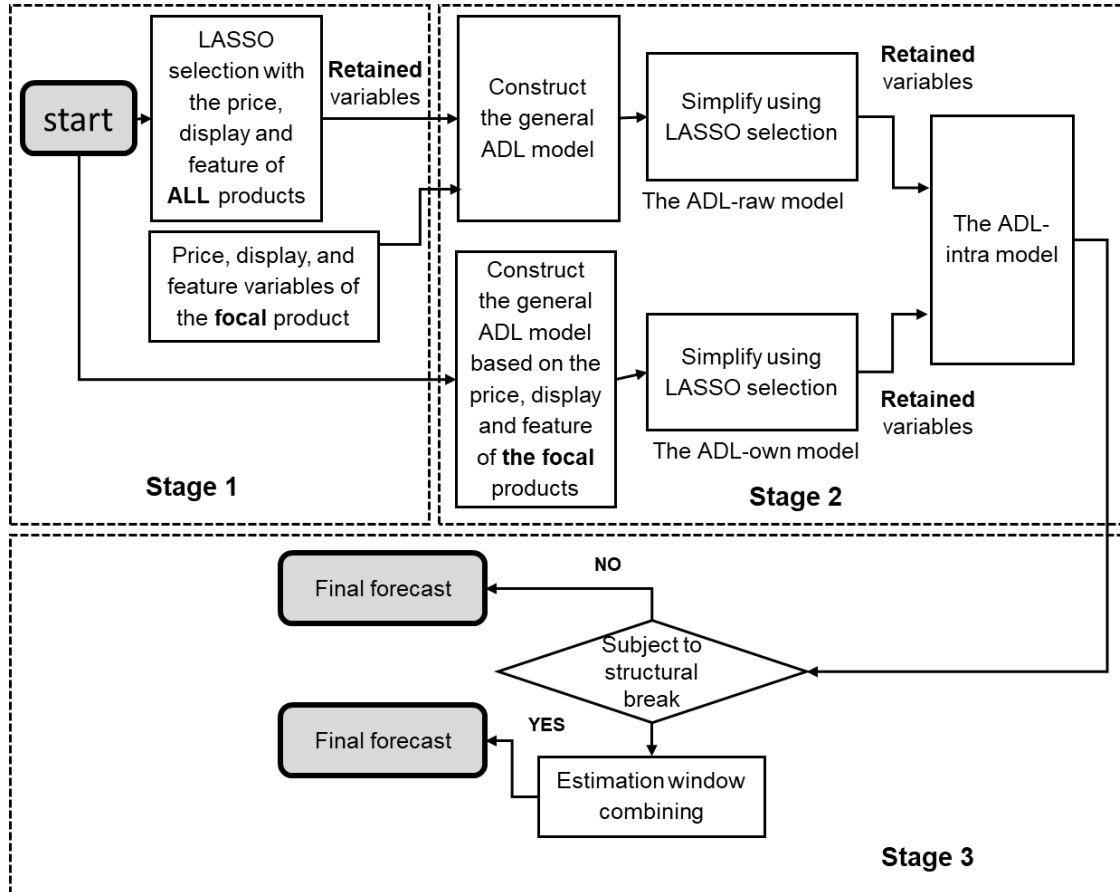
$$\begin{aligned} \ln(y_{0,t}) = & intercept + \tau * time + \sum_{j=1}^L \alpha_j \ln(y_{0,t-j}) + \sum_{j=0}^L \beta_{0,j} \ln(p_{0,t-j}) + \sum_{j=0}^L \gamma_{0,j} Feature_{0,t-j} \\ & + \sum_{j=0}^L \gamma_{0,j} Display_{0,t-j} + \sum_{d=1}^{12} \theta_d Four_week_dummy_d \\ & + \sum_{c=1}^9 \sum_{v=0}^1 \delta_{c,v} CalendarEvent_{c,t-v} + \varepsilon_t \end{aligned}$$

We also simplify this model using the LASSO procedure (we refer to this simplified model as the ADL-own model thereafter). We then incorporate the variables retained by the ADL-own model into the ADL-raw model (we refer the resulted model as the ADL-intra model). We include the variables

⁹ We include the following US calendar events including *Halloween*, *Thanksgiving*, *Christmas*, *New Year's Day*, *President's Day*, *Easter*, *Memorial Day*, the *4th of July*, and *Labour Day*.

in the ADL-own model because previous studies suggest that promotional variables of the focal variable are usually more important compared to variables of other products (Bucklin, Gupta, & Siddarth, 1998). We, therefore, reduce the probability of (wrongfully) discarding them at a cost of efficiency.

Figure 2. An illustration for the three-stages of the ADL-intra-EWC model



At the final stage, we integrate the ADL-intra model with the EWC method and the IC method respectively to take into account the structural break problem. We implement the EWC method and the IC method to the ADL-intra model only if the sequential Chow test indicates the existence of any structural break, and we keep the forecasts generated by the ADL-intra model as the final forecasts otherwise. We refer the models as the ADL-intra-EWC model and the ADL-intra-IC model respectively and we expect these models to generate more accurate forecasts by taking into account the structural break problem. Figure 2 provides a guide to implementing the ADL-intra-EWC model¹⁰.

7. The experimental design

¹⁰ The ADL-intra-IC model can be implemented analogously when the EWC method is replaced by the IC method if we confirm that the model is subject to structural break.

In this study, we initially evaluate the forecasting performance of the following models: 1) The Base-lift method¹¹; 2) The ADL-own model; 3) The ADL-intra model; 4) The ADL-intra-EWC model¹²; 5) The ADL-own-EWC model: similar to the ADL-intra-EWC model except that the ADL-intra model is replaced by the ADL-own model at the final stage; 6) The ADL-intra-IC model; 7) The ADL-own-IC model: similar to the ADL-intra-IC model except that the ADL-intra model is replaced by the ADL-own model at the final stage.

We evaluate the forecasting performance of these models with 18 rolling origins for robustness (Tashman, 2000). We specify the model with an estimation window of 160 weeks. For each rolling event, we move the estimation window two weeks forward and re-specify the model. We presume the value of the price and promotional information to be known, as it is part of the retailer's inventory plan, and we use the forecast value of the product sales when the forecast horizon is beyond one week. We generate one to H week-ahead forecasts, where H is 1, 4, and 8, to approximate the situation retailers face in practice. For the EWC method, we engage ten estimation windows with different lengths (e.g., for the initial estimation period $[1, 160]$, we estimate the model with ten estimation windows including $[1, 160]$, $[3, 160]$, and so forth, until $[19, 160]$), and generate ten sets of forecasts accordingly. We combine the ten sets of forecasts with equal weights. For the IC methods, we estimate the forecast bias as the average value of the sixteen most recent residuals and add the value equally to the forecasts of all the forecast horizons.

We evaluate the models with four error measures: the Mean Absolute Error (MAE), the symmetric Mean Absolute Percentage Error ($sMAPE$), the Mean Absolute Scaled Error ($MASE$) developed by Hyndman and Koehler (2006), and the Relative Average Mean Absolute Error ($RelAvgMAE$) developed by Davydenko and Fildes (2013). These error measures approximate the loss function of the retailer from different aspects. The error measures for S SKUs and K rolling events based on forecast horizon of 1 to H (i.e. $S = 1831$, $K = 18$, and $H=1, 4$ and 8) are as follows:

$$MAE(H) = \frac{1}{S} \frac{1}{K} \sum_{s=1}^S \sum_{k=1}^K MAE_{s,H,k}, \text{ where } MAE_{s,H,k} = \frac{1}{H} \sum_{h=1}^H |y_{s,h,k} - \hat{y}_{s,h,k}|$$

$$sMAPE(H) = \frac{1}{S} \frac{1}{K} \sum_{s=1}^S \sum_{k=1}^K sMAPE_{s,H,k}, \text{ where } sMAPE_{s,H,k} = \frac{1}{H} \sum_{h=1}^H \left| \frac{y_{s,h,k} - \hat{y}_{s,h,k}}{(y_{s,h,k} + \hat{y}_{s,h,k})/2} \right|$$

$$MASE(H) = \frac{1}{S} \sum_{s=1}^S \sum_{k=1}^K MASE(H)_k, \text{ where } MASE(H)_k = \frac{1}{S} \sum_{s=1}^S (|q_{s,H,k}|),$$

$$q_{s,H,k} = \frac{MAE_{s,H,k}^C}{\frac{1}{T_0 - 1} \sum_{t=2}^{T_0} |y_{s,t,k} - y_{s,t-1,k}|}$$

¹¹ More detailed descriptions can be found in Gür Ali et al. (2009) and Huang et al. (2014).

¹² We conduct the sequential Chow test and find the models for 99.89% of SKU's are subject to structural break.

$$AvgRelMAE(H) = \frac{1}{K} \sum_{k=1}^K \left(\prod_{s=1}^S RelMAE_{s,H,k} \right)^{\frac{1}{S}}, \text{ where } RelMAE_{s,H,k} = \frac{MAE_{s,H,k}^C}{MAE_{s,H,k}^B}$$

where $y_{s,h,k}$ and $\hat{y}_{s,h,k}$ are respectively the h^{th} actual value and forecast value of the forecast period for data series s based on the k^{th} rolling event. Before we transform the log values to levels, we adjust the final forecasts by adding one-half mean squared error (Cooper et al., 1999). T_0 is the total number of observations in the estimation window (i.e., $T_0 = 160$). $MAE_{s,H,k}^C$ and $MAE_{s,H,k}^B$ are the Mean Absolute Errors for the candidate model and the benchmark model respectively for data series s , with forecast horizon of H , for the k^{th} rolling event.

8. Results and discussion

In Table 2, we summarize the forecasting performance of the models across all the product categories. Table 3 shows the p-values of the Wilcoxon Sign Rank (WSR) test for the statistical significance of the difference between the models' forecasting performance. The WSR test is the non-parametric version of the traditional t -test without a normality assumption. We find the following from the analysis of the comparisons of forecasts from the different models:

- (i) The Base-lift model generates the least accurate forecasts.
- (ii) The ADL-intra model outperforms the ADL-own model, which demonstrates the value of competitive promotional information and corroborates with Huang et al. (2014).
- (iii) The ADL-own-EWC model significantly outperforms the ADL-own model.
- (iv) The ADL-own-IC model outperforms the ADL-own model for most of the error measures across lead times except for the volume dependent MAE error measure.
- (v) The ADL-intra-EWC model significantly outperforms the ADL-intra model.
- (vi) The ADL-intra-IC model outperforms the ADL-intra model for all the cases except for the MAE error measure.

Overall, The ADL-intra-EWC model and the ADL-intra-IC model generate the most accurate forecasts among the seven candidate models¹³.

¹³ The ADL-EWC-IC model in Table 3 will be discussed in later sections.

Table 2. The forecasting performance of the models for all forecast period

All forecast period, $H=8$									
Model/measure	MAE	Rank	SMAPE	Rank	MASE	Rank	AvgRelMAE	Rank	
Base-lift	22.92	8	47.0%	8	0.775	8	1.136	8	
ADL-own	15.76	6	40.8%	7	0.697	7	1.000	7	
ADL-intra	15.44	3	40.5%	4	0.695	5	0.991	4	
ADL-own-EWC	15.67	5	40.7%	5	0.696	6	0.996	6	
ADL-intra-EWC	15.35	2	40.4%	2	0.694	4	0.988	3	
ADL-own-IC	16.23	7	40.8%	6	0.694	3	0.997	5	
ADL-intra-IC	15.57	4	40.4%	3	0.692	2	0.988	2	
ADL-EWC-IC	15.29	1	40.4%	1	0.689	1	0.985	1	
All forecast period, $H=4$									
Model/measure	MAE	Rank	SMAPE	Rank	MASE	Rank	AvgRelMAE	Rank	
Base-lift	22.67	8	46.2%	8	0.762	8	1.106	8	
ADL-own	15.63	6	40.5%	7	0.690	7	1.000	7	
ADL-intra	15.16	3	40.1%	4	0.686	5	0.989	4	
ADL-own-EWC	15.55	5	40.3%	6	0.688	6	0.995	6	
ADL-intra-EWC	15.08	2	40.0%	3	0.685	4	0.985	3	
ADL-own-IC	15.94	7	40.3%	5	0.684	3	0.993	5	
ADL-intra-IC	15.19	4	39.9%	2	0.681	2	0.982	2	
ADL-EWC-IC	15.00	1	39.9%	1	0.679	1	0.980	1	
All forecast period, $H=1$									
Model/measure	MAE	Rank	SMAPE	Rank	MASE	Rank	AvgRelMAE	Rank	
Base-lift	24.99	8	45.4%	8	0.762	8	1.026	8	
ADL-own	16.66	6	39.9%	7	0.689	7	1.000	7	
ADL-intra	15.66	4	39.4%	4	0.686	5	0.980	4	
ADL-own-EWC	16.58	5	39.7%	6	0.686	6	0.994	6	
ADL-intra-EWC	15.59	2	39.3%	3	0.684	4	0.976	3	
ADL-own-IC	17.01	7	39.6%	5	0.681	3	0.982	5	
ADL-intra-IC	15.60	3	39.2%	1	0.678	1	0.966	2	
ADL-EWC-IC	15.50	1	39.2%	2	0.679	2	0.965	1	

Table 3. The p-values of the Wilcoxon Sign Rank (WSR) test

Candidate model		MAE			SMAPE			MASE			RelMAE		
Benchmark		$H=1$	$H=4$	$H=8$	$H=1$	$H=4$	$H=8$	$H=1$	$H=4$	$H=8$	$H=1$	$H=4$	$H=8$
ADL-intra	ADL-intra-EWC	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ADL-intra	ADL-intra-IC	0.07	0.73	0.00	0.04	0.43	0.09	0.08	0.89	0.02	0.16	0.65	0.00
ADL-own	ADL-own-EWC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ADL-own	ADL-own-IC	0.36	0.31	0.00	0.04	0.69	0.04	0.14	0.68	0.01	0.29	0.20	0.00
ADL-own	ADL-intra	0.02	0.11	0.26	0.00	0.01	0.04	0.02	0.14	0.29	0.04	0.55	0.91
ADL-own	Base-lift	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ADL-intra	ADL-EWC-IC	0.07	0.17	0.33	0.04	0.17	0.29	0.11	0.09	0.46	0.36	0.90	0.02

We also investigate the models' forecasting performance for the time period depending on whether or not the focal product is being promoted as the sales for these two periods tend to exhibit different

levels of variations¹⁴. Table 4 shows the forecasting performance of the models for the non-promoted period and the promoted forecast period respectively for one to eight-week forecast horizon¹⁵. The results are similar compared to those in Table 2. Of the many detailed comparisons possible, the following seem particularly important:

- (i) Models that include competitive information (e.g., ADL-intra) outperform those that only include focal product information (e.g., ADL-own).
- (ii) Models that include the possibility of structural breaks generally outperform those that neglect the possibility. The only exception is, for the promoted periods, the models which integrate intercept corrections do not effectively outperform their counterparts. This may be due to the high volumes and high variations of the product sales when the focal product is being promoted which submerge the value of the bias correction, a point we return to later.

Table 4. The forecasting performance of the models for the promoted and the non-promoted period

H=8	Non-Promoted period				Promoted period			
Model/measure	MAE	SMAPE	MASE	AvgRelMAE	MAE	SMAPE	MASE	AvgRelMAE
Base-lift	9.64	41.7%	0.588	1.005	84.46	82.1%	2.186	1.504
ADL-own	9.36	40.3%	0.582	1	49.49	49.4%	1.653	1
ADL-intra	9.13	40.1%	0.582	0.996	48.78	48.0%	1.63	0.972
ADL-own-EWC	9.33	40.2%	0.581	0.996	49.23	49.3%	1.652	0.995
ADL-intra-EWC	9.1	40.0%	0.581	0.994	48.34	47.9%	1.626	0.965
ADL-own-IC	9.23	40.2%	0.575	0.995	51.54	49.7%	1.673	1.017
ADL-intra-IC	9.03	40.0%	0.577	0.992	49.71	48.3%	1.647	0.988
ADL-EWC-IC	9.03	40.0%	0.577	0.992	48.34	47.9%	1.626	0.965

In Table 5, we compare the forecasting performance of the ADL-intra-EWC model and the ADL-inter-IC model to the ADL-intra model for each individual product category based on the MASE for one to eight-week forecast horizon. We focus on the ADL-intra-EWC model and the ADL-inter-IC model because they have the best forecasting performance overall and the ADL-intra model is their counterpart model which overlooks the issue of structural break. We show the forecasts based on the MASE for one to eight-week horizon for simplicity and the results for other measures and horizons are similar. The ADL-intra-EWC model and the ADL-intra-IC model outperforms the ADL-intra model for 20 and 19 product categories respectively (out of 28 categories). The ADL-intra-EWC model and the ADL-inter-IC model do not outperform the ADL-intra model for every product category due to the heterogeneity of the data characteristics across different product categories (e.g., Ma et al., 2016). The ADL-intra-EWC model or the ADL-intra-IC model reduce the MASE by 1.97% and 2.9% respectively for the 20% of categories (i.e., 6 categories) where they achieve their maximum

¹⁴ We refer these two periods as the promoted period and non-promoted period respectively.

¹⁵ The results for other forecasting horizons are similar and are not shown here for simplicity.

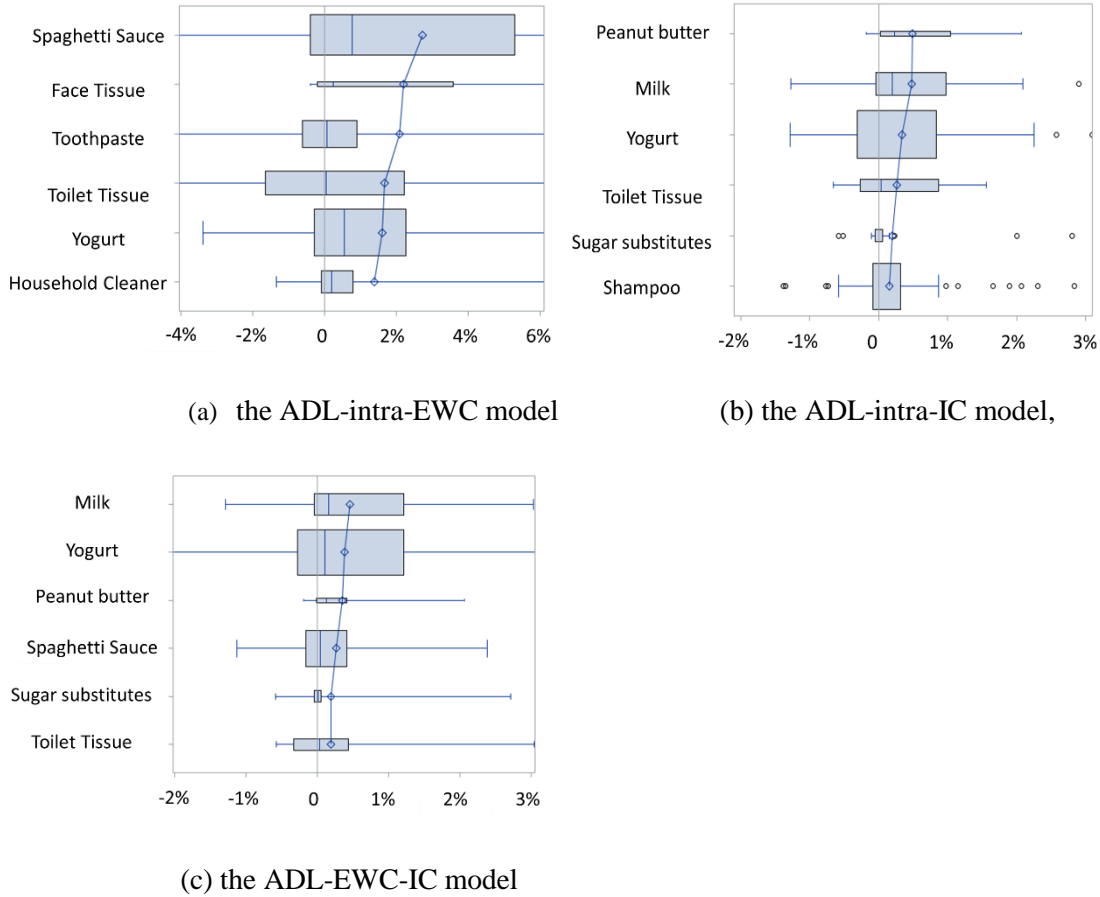
improvement. Figure 3 shows further details using boxplots for those product categories. In Figure 3(a) (i.e., the boxplot in the upper-left corner) and Figure 3(b) (i.e., the boxplot in the upper-right corner), positive values indicate the percentage improvements by the ADL-intra-EWC model and by the ADL-intra-IC model compared to the ADL-intra model.

Table 5. The relative forecasting performance of the ADL-intra-EWC model and the ADL-intra-IC model compared to the ADL-intra model for each product category for the MASE for one to eight-week forecast horizon¹⁶

Category/MASE	ADL-intra-EWC	ADL-intra-IC	Category/MASE	ADL-intra-EWC	ADL-intra-IC
Beer	99.90%	100.40%	Mayonnaise	99.96%	99.58%
Blades	99.76%	97.35%	Milk	99.06%	94.14%
Carbonated Beverages	99.56%	99.69%	Mustard & Ketchup	99.36%	100.95%
Cigarette	99.85%	98.77%	Peanut butter	100.16%	94.96%
Coffee	100.06%	99.20%	Photo	99.03%	99.42%
Cold Cereal	99.70%	101.60%	Salty snacks	100.03%	99.77%
Deodorant	100.00%	98.47%	Shampoo	99.74%	98.52%
Face Tissue	98.21%	100.60%	Soup	99.02%	103.14%
Frozen Dinner	100.71%	100.20%	Spaghetti sauce	98.40%	98.60%
Frozen pizza	101.67%	101.55%	Sugar substitutes	99.67%	96.51%
Household Cleaner	98.78%	99.83%	Toilet Tissue	99.66%	98.62%
Hotdog	100.43%	103.25%	Toothbrush	100.13%	101.27%
Laundry Detergent	99.54%	99.58%	Toothpaste	98.27%	100.25%
Margarine/Butter	100.42%	100.71%	Yogurt	98.25%	95.52%

¹⁶ Percentage values lower than one indicate that the ADL-intra model is outperformed for that product category.

Figure 3. Compare three models with the ADL-intra model for six product categories: results at SKU level, for the MASE, and for one to eight-week forecast horizon¹⁷.



The results in Table 4 indicates that the ADL-intra-IC model has the best forecasting performance for the non-promoted period but only has a moderate performance for the promoted period. A possible explanation is that the estimated bias used for the correction gets submerged by high variations in the high product sales. This allows us to complement the ADL-intra-IC model for the promoted period. As the ADL-intra-EWC model has the best performance for the promoted period, we forge a combined model between these two models, named as the ADL-EWC-IC model. The ADL-EWC-IC model will be identical to the ADL-intra-EWC model for the promoted period and the ADL-intra-IC model for the non-promoted period. Table 2 also shows the forecasting performance by the ADL-EWC-IC model. The results indicate that the ADL-EWC-IC model generates the most accurate forecasts across all the models. Table 4 also includes the performance of the ADL-EWC-IC model for the promoted and non-promoted forecast periods. The ADL-EWC-IC model outperforms the ADL-intra model for 21 (out of 28) product categories. It reduces the MASE by 3.3% for the 20% of categories (i.e., 6 categories) where it achieves the maximum improvement. Figure 3(c) (i.e., the

¹⁷ The widths of the boxplots are determined by the number of SKU's in each product category.

boxplot in the bottom-right corner) shows the percentage improvements of the ADL-EWC-IC model compared to the ADL-intra model.

9. Exploring the determinants of the forecasting improvement

The results show that our proposed models generate more accurate forecasts especially for some product categories (e.g., Yogurt, Milk, Toilet Tissue etc.). We further explore the determinants of the improvement of the forecasting performance of our proposed models at SKU level. This provides insights into what types of SKUs may benefit most from the proposed models. We consider the following types of potential determinants: 1) basic statistical measures for both the prices and sales including the average, standard deviation, skewness, range, kurtosis, and coefficient of variation; 2) the frequency of the feature and display promotions for each SKU. 3) more advanced statistical measures which capture the characteristics of the data series designed by Fildes (1992). For example, we measure the proportion of outliers for the sales of each SKU. The value of the sales for product i will be identified as an outlier if $\Delta y_i < Q_1 - 1.5 * (Q_3 - Q_1)$ or $\Delta y_i > Q_3 + 1.5 * (Q_3 - Q_1)$, where Δy_i is the differenced value of the sales for product i . Q_1 and Q_3 are the first and third quantiles of Δy_i . We measure randomness by regressing $y'_{i,t}$ on T , $y'_{i,t-1}$, $y'_{i,t-2}$, and $y'_{i,t-3}$, where $y'_{i,t}$ is the sales value for product i at week t given that the outliers are removed and T is the time trend. The fitness of this autoregressive model (e.g., the R square) approximates the systematic variation in the sales data series which could be captured by simple models. Lastly, we measure the linear trend for the sales of the SKU as the absolute value of the correlation between $y'_{i,t}$ and the time trend.

We develop five orthogonal factors out of the fourteen explanatory variables above to mitigate the issue of multicollinearity¹⁸. Table 6 shows the correlation between the original fourteen explanatory variables and the constructed factors¹⁹. We interpret factor 1 as “Outliers and general variations”, factor 2 as “Sales level and variation”, factor 3 as “Central tendency of sales”, factor 4 as “Price level and variation”, and factor 5 as “Randomness and growth”. We then regress the percentage improvement by the models based on these 5 factors at SKU level. For robustness, we construct the model with and without dummy variables for product categories.

Table 6. The pattern of the factors

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Proportion of outliers	0.855				
Coefficient of variation (price)	0.759				
Coefficient of variation (sales)	0.714				

¹⁸ We retain 90% of the variations with five factors.

¹⁹ In Table 6, we omit all small values for simplicity.

	Frequency of Feature	0.703		
1	Standard deviation of sales	0.964		
2	Range of sales	0.929		
3	Average sales	0.807		
4	Frequency of Display	0.281		
5	Kurtosis of sales		0.973	
6	Skewness of sales		0.881	
7	Standard deviation of price		0.987	
8	Average price		0.831	
9	Randomness			0.992
10	Trend			0.703
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				

Table 7 The determinants of improvement (MASE) by the candidate model compared to their counterparts*

Horizon = 8	ADL-intra-EWC		ADL-own-EWC		ADL-intra-IC		ADL-own-IC		ADL-EWC-IC		IC versus EWC	
Parameter/estimate and p-values	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Outliers and general variations	0.1	0.319	0.1	0.321	-1	0.000	-1.4	0.000	-1.1	0.000	-0.8	0.000
Sales level and variation	0.1	0.134	0.2	0.085	-0.2	0.277	-1	0.000	-0.4	0.082	0.0	0.898
Central tendency of sales	-0.1	0.508	-0.1	0.530	-0.7	0.001	-0.8	0.001	-0.7	0.001	-0.3	0.059
Price level and variation	-0.1	0.170	-0.2	0.121	0	0.824	-0.1	0.761	0.2	0.398	0.1	0.352
Randomness and growth	0.4	0.000	0.4	0.000	0.6	0.008	0.7	0.003	0.2	0.461	0.6	0.000
Intercept	0.3	0.001	0.3	0.001	-0.2	0.234	-0.4	0.094	-0.6	0.004	0.1	0.304
Model with 5 factors and category dummy variables												
Horizon = 8	ADL-intra-EWC		ADL-own-EWC		ADL-intra-IC		ADL-own-IC		ADL-EWC-IC		IC versus EWC	
Parameter/estimate and p-values	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Outliers and general variations	0.2	0.085	0.4	0.013	-0.5	0.155	-0.7	0.075	-0.7	0.028	-0.3	0.137
Sales level and variation	0.1	0.150	0.2	0.054	-0.1	0.539	-0.9	0.000	-0.3	0.210	0.0	0.824
Central tendency of sales	0	0.679	0	0.851	-0.5	0.044	-0.5	0.047	-0.5	0.027	-0.1	0.559
Price level and variation	-0.1	0.370	-0.3	0.066	-0.1	0.795	-0.3	0.367	0.1	0.841	0.2	0.376
Randomness and growth	0.3	0.001	0.4	0.000	0.4	0.053	0.5	0.055	0.1	0.676	0.3	0.024
Intercept	1.5	0.001	1.6	0.001	2.6	0.015	4.2	0.001	1.2	0.263	1.9	0.008

*the estimates are all multiplied
by 100

Table 7 reports the estimated parameters of the models. The dependent variables are the percentage improvement of the candidate models compared with their counterparts which overlook the structural break problem for the MASE and when the horizon is one to eight-week ahead at SKU level. The results are consistent across other error measures and forecast horizons. For the ADL-intra-EWC model over the ADL-intra model without product category dummies the estimate of the parameter “Randomness and growth” is positive (e.g., 0.4) and statistically significant (e.g., p-value<0.001): similarly, for the extended model with category dummy variables included. This result holds true also for the models which only include the focal marketing variables. This suggests that our proposed models tend to be more advantageous for the SKUs which are difficult to forecast and exhibit a trend in sales, possibly because the SKUs with higher levels of ‘randomness and trend’ are more heavily associated with the structural break problem and forecast bias. Second, the IC related models including the ADL-intra-IC model, the ADL-own-IC model, and the ADL-EWC-IC model tend to have disadvantages for the SKUs with a high proportion of outliers and for the SKUs with the high central tendency of sales. This may indicate that the ‘intercept correction’ for the bias can be submerged by high sales peaks which are usually ‘outliers’ and caused by promotions. We also investigate the determinants of the advantages of the ADL-intra-IC model over the ADL-intra-EWC model (e.g., the last two columns in Table 7). The results suggest that the ADL-intra-IC model tend to be more advantageous for the SKUs which are difficult to forecast and exhibit randomness and trend in sales. All the results here indicate that we may pre-test these features for each SKU and then determine the optimal sales forecasting method specifically for that SKU.

Table 8. Forecasting performance regarding percentage reductions in various error measures for the one to eight-week forecast horizon

Proposed model	Benchmark	Percentage of increase/decrease			
		MAE	SMAPE	MASE	AvgRelMAE
ADL-intra-EWC	ADL-intra	-0.57%	-0.22%	-0.18%	-0.33%
ADL-intra-IC	ADL-intra	0.86%	-0.18%	-0.45%	-0.36%
ADL-EWC-IC	ADL-intra	-0.93%	-0.27%	-0.85%	-0.66%
ADL-own-EWC	ADL-own	-0.52%	-0.31%	-0.17%	-0.42%
ADL-own-IC	ADL-own	3.00%	-0.15%	-0.52%	-0.32%
ADL-intra	ADL-own	-2.02%	-0.75%	-0.30%	-0.86%
ADL-own-EWC	Base-lift	-31.61%	-13.40%	-10.22%	-12.36%
ADL-own-IC	Base-lift	-29.20%	-13.26%	-10.53%	-12.27%
ADL-intra-EWC	Base-lift	-33.03%	-13.97%	-10.49%	-13.04%
ADL-intra-IC	Base-lift	-32.07%	-13.94%	-10.74%	-13.06%
ADL-EWC-IC	Base-lift	-33.27%	-14.02%	-11.09%	-13.32%

10. Conclusions, limitations and future research

Grocery retailers need to effectively manage their inventory and, to achieve that, they rely on forecasting models and welcome new approaches that will enable them to improve their current practices. Related studies focus on incorporating more information (e.g., Gür Ali et al., 2009; Huang et al., 2014; Ma et al., 2016). These studies all assume that the effect of the marketing activities such as price reductions and feature and display promotions remain unchanged over time. This assumption may not hold because of the impact of external factors including the change in economic conditions, the change of the consumer taste, and new competition entry etc. The data on these factors are not always available to incorporate into the model. Or, we do not actually know which of these external factors are actually causing the structural break. As a result, the conventional models with all the available data used in the model building may be subject to the problem of structural break and potentially generate biased and less accurate forecasts.

Our research focuses on how to mitigate the problem using data on the marketing activities which retailers typically have control over. We propose a set of models which take into account the potential forecast bias caused by structural breaks. The ADL-intra-EWC model generates forecasts which are the combination of various sets of forecasts by the ADL-intra model with different estimation windows under the condition where structural breaks are detected. It tries to obtain an effective trade-off between the forecast bias and the forecast error variance. The ADL-intra-IC model tries to offset the potential forecast bias by adding the estimate of the forecast bias back to the error term at a cost of inflated forecast error variance when structural breaks are detected. In the retailer context, the data at SKU level exhibit very different characteristics across different product categories and usually exhibit high levels of variations. Based on our empirical results, we find that these models outperform the ADL-intra model across all the 28 product categories. Table 8 shows the percentage reductions of (OrderDynamics, 2015) week forecast horizon²⁰. For example, the ADL-intra-EWC model reduces the MASE of the ADL-intra model by 0.18% and reduces the MASE of the Base-lift model by 10.49%. The ADL-intra-IC model reduces the MASE of the ADL-intra model by 0.45% and reduces the MASE of the Base-lift model by 10.74%. The ADL-EWC-IC model reduces the MASE of the ADL-intra model by 0.85% and reduces the MASE of the Base-lift model by 11.09%. The improved forecasting accuracy for product sales substantially contributes to retailers' profit, considering the economic scale of their business (Cooper et al., 1999; OrderDynamics, 2015).

More important than these average reductions, at the category level, these models have superior forecasting performance compared to the ADL-intra model for most of the product categories.

²⁰ The results are similar for other forecast horizons.

We observe that the ADL-intra-EWC model has the best performance for the promoted forecast period while the ADL-intra-IC model dominates the non-promoted forecast period. We, therefore, forge a model combining the ADL-intra-EWC model and the ADL-intra-IC model based on whether or not the focal product is being promoted. The so-called ADL-EWC-IC model thus generates the most accurate forecasts across all the candidate models. The ADL-EWC-IC model has superior forecasting performances compared to the ADL-intra model for 21 out of 28 product categories with an overall improvement of 0.85% compared to the ADL-intra model and 11.09% compared to the Base-lift model for the MASE for the forecast horizon of one to eight-week ahead.

In this study, we also evaluate the forecasting performance of the ADL-own-EWC model and the ADL-own-IC model. These methods are especially valuable to manufacturers since, under certain circumstances, competitive promotional information may not be available to them (Ali & Boylan, 2011; Ali, Babai, Boylan, & Syntetos, 2017). In our experiment, the ADL-own -EWC model and the ADL-own -IC model both outperform the ADL-own model across all the product categories. Table 7 shows the percentage reductions of various error measures by these models compared to different benchmarks. For example, the ADL-own-EWC model reduces the MASE of the ADL-own model by 0.17% and reduces the SMAPE of the Base-lift model by 10.22%. The ADL-own-IC model reduces the MASE of the ADL-own model by 0.52% and reduces the MASE of the Base-lift model by 10.53%.

We also explore the relationship between the relative advantage of the proposed models and the data characteristics of the product SKU. We find that the models with the estimation window combining method (e.g., the ADL-intra-EWC model and the ADL-own-EWC model) have better forecasting performances compared to their counterparts for the SKU's with high randomness and trend, while the models with intercept corrections (e.g., the ADL-intra-IC model, the ADL-own-IC model, and the ADL-EWC-IC model) tend to have more advantages compared to their counterparts for the SKU's with high randomness and trend, with a low proportion of outliers and low level of general variations, and with a low level of sales central tendency.

The approach that we propose in this study is new to the area of retailer product sales forecasting but we have also identified some areas where we feel further improvements could be beneficial. For example, for the EWC method, we combine five sets of forecasts based on ten different estimation windows using equal weights. The forecasting performance may potentially be improved by changing the number of the estimation windows, by changing the minimum length of the estimation windows, and by exploring alternative forecasting combination schemes (e.g., based on k-fold evaluation). For the IC method, Clements and Hendry (1999) summarize various correction schemes each of which

may have different effects on the trade-off between the bias and the error variance²¹. Furthermore, Ma et al. (2016) propose models which integrate both the intra- and the inter-category promotional information. Thus, we may further investigate how we can improve the forecasting performance with both the intra- and the inter-category promotional information while taking into account the structural break problem. A method alternative to the ADL-intra-EWC method and the ADL-intra-IC method is directly modeling the changing process of the effect of the marketing activities into the model through, for example, random coefficient models so that the structural break may potentially be mitigated even when the influencing factors are not observed. A potential disadvantage of this method is that we need to make strong assumptions of how the effect of the marketing activities actually changes. Foekens, Leeftang, and Wittink (1999) proposed market response models with time-varying parameters but the model was not for forecasting purposes. In summary, the models we proposed in this study produce consistently more accurate forecasts and they also take into account the practical requirements of retail forecasting in that they are intuitive, they can be developed and operated automatically and also use readily available data on marketing activities.

Acknowledgement

We thank the IRI company for making the data available. All the analysis and findings in this paper based on the IRI dataset are by the authors and not by the IRI company.

Reference:

- Ali, M., & Boylan, J. (2011). Feasibility principles for Downstream Demand Inference in supply chains. *Journal of the Operational Research Society*, 62.
- Ali, M. M., Babai, M. Z., Boylan, J. E., & Syntetos, A. A. (2017). Supply chain forecasting when information is not shared. *European Journal of Operational Research*, 260(3), 984-994. doi: <https://doi.org/10.1016/j.ejor.2016.11.046>
- Allen, P. G., & Fildes, R. (2001). Econometric forecasting. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Boston: Kluwer Academic Publishers.
- Andrews, R. L., Currim, I. S., Leeftang, P., & Lim, J. (2008). Estimating the SCAN*PRO model of store sales: HB, FM or just OLS? *international Journal of research in marketing*, 25(1), 22-33.
- Ang, A., & Bekaert, G. (2002). Regime Switches in Interest Rates. *Journal of Business & Economic Statistics*, 20(2), 163-182.
- Armstrong, J. S. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers.

²¹ For example, one of the alternative is to first make adjustments to the one-step-ahead forecast, and then calculate the two-step-ahead forecast based on the value of the one-step-ahead forecast which has adjusted, and so forth.

- Blattberg, R. C., Briesch, R., & Fox, E. J. (1995). How promotions work? *Marketing Science*, 14(3).
- Bronnenberg, B. J., Kruger, M. W., & Mela, C. F. (2008). The IRI Marketing Data Set. *Marketing Science*, 27(4), pp. 745–748.
- Bucklin, R. E., Gupta, S., & Siddarth, S. (1998). Determining Segmentation in Sales Response across Consumer Purchase Behaviors. *Journal of Marketing Research*, 35(2), 189-197. doi: 10.2307/3151847
- Castle, J. L., Doornik, J. A., & Hendry, D. F. (2008). Model Selection when there are Multiple Breaks. *Working paper No. 407, Economics Department, University of Oxford*.
- Chevillon, G. (2016). Multistep forecasting in the presence of location shifts. *International Journal of Forecasting*, 32(1), 121-137.
- Clark, T. E., & McCracken, M. W. (2007). Forecasting with Small Macroeconomic VARs in the Presence of Instabilities *Finance and Economics Discussion Series*
- Federal Reserve Board, Washington, D.C.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559-583.
- Clements, M., & Hendry, D. (1998). *Forecasting Economic Time Series*: Cambridge University Press.
- Clements, M. B., & Hendry, D. F. (1994). Towards a theory of economic forecasting. In C. P. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*: Oxford University Press.
- Clements, M. P., & Hendry, D. F. (1999). *Forecasting non-stationary economic time series*. London: The MIT Press.
- Cooper, L., Baron, P., Levy, W., Swisher, M., & Gogos, P. (1999). Promocast": a New Forecasting Method for Promotion Planning. *Marketing Science*, 18(3), 301-316.
- Cooper, L. G., & Giuffrida, G. (2000). Turning Datamining into a Management Science Tool: New Algorithms and Empirical Results. *Management Science*, 46(2), 249.
- Corsten, D., & Gruen, T. (2003). Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management*, 31(12).
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: the case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3).
- Dekimpe, M., Hanssens, D. M., & Silva-Risso, J. M. (1999). Long-run effects of price promotions in scanner markets. *Journal of Econometrics*, 89(1/2), 261-291.
- Dekker, M., van Donselaar, K., & Ouwehand, P. (2004). How to use aggregation and combined forecasting to improve seasonal demand forecasts. *International Journal of Production Economics*, 90(2), 151-167.
- Epprecht, C., Guegan, D., & Veiga, Á. (2013). Comparing variable selection techniques for linear regression: LASSO and Autometrics: Université Panthéon-Sorbonne (Paris 1), Centre d'Economie de la Sorbonne.
- Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *Journal of Royal Statistical Society*, 70(Series B), 849–911.
- Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, 8, 81-98.
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), 3-23.

- Fildes, R., Nikolopoulos, K., Crone, S., & Syntetos, A. (2008). Forecasting and operational research: A review. *Journal of the Operational Research Society*, 59.
- Fildes, R., & Stekler, H. (2002). The state of macroeconomic forecasting. *Journal of Macroeconomics*, 24(4), 435-468. doi: Pii S0164-0704(02)00055-1
- Foekens, E. W., Leeflang, P., & Wittink, D. R. (1999). Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometrics*, 89(1-2), 249-268.
- Goodwin, P. (2002). Integrating management judgment and statistical methods to improve short-term forecasts. *Omega*, 30(2), 127-135. doi: Doi: 10.1016/s0305-0483(01)00062-7
- Gür Ali, Ö., SayIn, S., van Woensel, T., & Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10).
- Hendry, D. F. (2018). Deciding between alternative approaches in macroeconomics. *International Journal of Forecasting*, 34(1), 119-135. doi: <https://doi.org/10.1016/j.ijforecast.2017.09.003>
- Houston, F. S., & Weiss, D. L. (1975). Cumulative advertising effects: the role of serial correlation. *Decision Sciences*, 6(3), 471-481. doi: 10.1111/j.1540-5915.1975.tb01036.x
- Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting FMCG retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2), 738-748.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679-688.
- Jose, V. R. R., & Winkler, R. L. (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, 24(1), 163-169.
- Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, 23(3), 377-390. doi: <https://doi.org/10.1016/j.ijforecast.2007.02.006>
- Little, J. D. C. (1966). A Model of Adaptive Control of Promotional Spending. *Operations research*, 14(6).
- Loeb, W. (2015). Unrelenting Competition: The Biggest Retail Story of 2015, 2016
- Ma, S., & Fildes, R. (2017). A retail store SKU promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research*, 260(2), 680-692.
- Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-category promotional information. *European Journal of Operational Research*, 249(1), 245-257.
- Mace, S., & Neslin, S. A. (2004). The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research*, XLI, 339-350.
- Mahajan, V., Bretschneider, S. I., & Bradford, J. W. (1980). Feedback Approaches to Modeling Structural Shifts in Market Response. *Journal of Marketing*, 44, 71-80.
- Martin, R., & Kolassa, S. (2009). *Challenges of Automated Forecasting in Retail*. Paper presented at the International Symposium on Forecasting, Hong Kong.
- Meeran, S., Jahanbin, S., Goodwin, P., & Quariguasi Frota Neto, J. (2017). When do changes in consumer preferences make forecasts from choice-based conjoint models unreliable? *European Journal of Operational Research*, 258(2), 512-524. doi: <https://doi.org/10.1016/j.ejor.2016.08.047>
- Moinpour, R., McCullough, J. M., & MacLachlan, D. L. (1976). Time Changes in Perception: A Longitudinal Application of Multidimensional Scaling. *Journal of marketing research*, 13(3), 245-253.

- Monroe, K. B., & Gultinan, J. P. (1975). A Path-Analytic Exploration of Retail Patronage Influences. *The Journal of Consumer Research*, 2(1), 19-28.
- Morrison, D. G. (1966). Interpurchase Time and Brand Loyalty. *Journal of Marketing Research*, 3.
- Myers, J. G. (1971). The Sensitivity of Dynamic Time-Path Typologies. *Journal of marketing research*, 8(4), 472-479.
- Myers, J. G., & Nicosia, F. M. (1970). Time-Path Types: From Static to Dynamic Typologies. *Management Science*, 16(10), B584-B596.
- Nijs, V. R., Dekimpe, M. G., Steenkamps, J.-B. E. M., & Hanssens, D. M. (2001). The Category-Demand Effects of Price Promotions. *Marketing Science*, 20(1), 1-22.
- Nikolopoulos, K. (2010). Forecasting with quantitative methods: the impact of special events in time series. *Applied Economics*, 42, 947-955.
- OrderDynamics. (2015). Retailers and the Ghost Economy: The Haunting of Returns. http://engage.dynamicaaction.com/WS-2015-06-IHL-Ghost-Economy-Haunting-of>Returns-AR_LP.html.
- Ouyang, Y. (2007). The effect of information sharing on supply chain stability and the bullwhip effect. *European Journal of Operational Research*, 182, 1107-1121.
- Pesaran, H. M., & Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129(1-2), 183-217. doi: DOI: 10.1016/j.jeconom.2004.09.007
- Pesaran, H. M., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137, 134-161.
- Pesaran, M. H., & Pick, A. (2011). Forecast Combination Across Estimation Windows. *Journal of Business & Economic Statistics*, 29(2), 307-318. doi: 10.1198/jbes.2010.09018
- Pesaran, M. H., Schuermann, T., & Smith, L. V. (2009). Forecasting Economic and Financial Variables with Global VARs. *International Journal of Forecasting*, 25, 642-675.
- Petropoulos, F., Fildes, R., & Goodwin, P. (2016). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? *European Journal of Operational Research*, 249(3), 842-852. doi: <https://doi.org/10.1016/j.ejor.2015.06.002>
- Petropoulos, F., Makridakis, S., Assimakopoulos, V., & Nikolopoulos, K. (2014). 'Horses for Courses' in demand forecasting. *European Journal of Operational Research*, 237(1), 152-163. doi: <http://dx.doi.org/10.1016/j.ejor.2014.02.036>
- Sawaya, W. J., & Giauque, W. C. (1986). *Production and operations management*: Harcourt Brace Jovanovich.
- Sodhi, M. S., & Tang, C. S. (2011). The incremental bullwhip effect of operational deviations in an arborescent supply chain with requirements planning. *European Journal of Operational Research*, 215, 374-382.
- Song, H., & Witt, S. F. (2003). Tourism Forecasting: The General-to-Specific Approach. *Journal of Travel Research*, 42, 65-74.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review *International Journal of Forecasting*, 16(4).
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.
- Trusov, M., Bodapati, A. V., & Cooper, L. G. (2006). Retailer Promotion Planning: Improving Forecasting Accuracy And Interpretability. *Journal of Interactive Marketing*, 20(3-4), 71-81.
- Van Heerde, H. J., Gupta, S., & Wittink, D. R. (2003). Is 75% of the Sales Promotion Bump Due to Brand Switching? No, Only 33% Is. *Journal of Marketing Research*, XL, 481-491.

- 1 Van Heerde, H. J., Srinivasan, S., & Dekimpe, M. G. (2008). Decomposing the Demand for a
2 Pioneering Innovation. *Working paper, University of Waikato, Department of*
3 *Marketing*.
- 4 Wedel, M., & Zhang, J. (2004). Analyzing brand competition across subcategories. *Journal*
5 *of Marketing Research*, 41(4), 448-456.
- 6 Wichern, D. W., & Jones, R. H. (1977). Assessing the Impact of Market Disturbances Using
7 Intervention Analysis. *Management Science*, 24(3), 329-337.
- 8 Wildt, A. R. (1976). The empirical investigation of time dependent parameter variation in
9 marketing models. In E. proceedings (Ed.), *American Marketing Association* (pp.
10 466-472).
- 11 Wildt, A. R., & Winer, R. S. (1983). Modeling and Estimation in Changing Market
12 Environments. *The Journal of Business*, 56(3).
- 13 Winer, R. S. (1979). An Analysis of the Time-varying Effects of Advertising: The Case of
14 Lydia Pinkham. *The Journal of Business*, 52(4).
- 15 Wittink, D., Addona, M., Hawkes, W., & Porter, J. (1988). *SCAN*PRO: the estimation,*
16 *validation and use of promotional effects based on scanner data*. Internal paper.
17 Internal Paper. Cornell University.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Supplementary Material

[Click here to download Supplementary Material: Huang Fildes Soopramanien 2018 SUPPLEMENTARY MATERIALS.docx](#)