**Forecasting Retailer Product Sales in the Presence of Structural Breaks**

Tao Huang, Robert Fildes, Didier Soopramanien

Abstract

Retailers need accurate sales forecasts for their inventory management. In this study we propose effective methods to generate more accurate forecasts by taking into account the issue of structural breaks and forecast bias caused by unobservable influencing factors. We propose three stages models based on the Autoregressive Distributed Lag (ADL) model with intercept correction and estimation window combining. With the intercept correction technique we try to offset the forecast bias caused by the structural break. With the estimation window combining technique we try to improve forecasting accuracy with a better trade-off between forecast bias and forecast error variance. We evaluate our models for products in a wide range of product categories and we found the proposed new models have the best forecasting performance.

Key words:

Sales Forecasting, Marketing analytics, Promotion

**Section 1: Introduction**

Retailers deals with out-of-stock and over stock by more accurate forecasts.

Retailers have been struggling with the situations of out-of-stock and over-stock for years. When a product is out-of-stock, retailers not only lose profits but also may lose the customers forever. Previous studies show that customers whom were once believed to either purchase alternative products or postpone their purchases when their preferred products are out of stock are actually more likely to switch to other stores and never come back ([Corsten and Gruen 2003](#_ENREF_16)). In practice, retailers try may deliberately increasing the inventory level (i.e. to over-stock) to avoid the out-of-stock condition, which however significantly raises inventory costs and reduces profits (Cooper, Baron et al. 1999). Under such a circumstance, retailers need to balance the loss due to running out-of-stock and the cost of higher inventory level. One of the keys to resolve this cost and service level dilemma is to generate accurate forecasts for the product sales (Corsten and Gruen 2003).

Accurate forecasts are difficult to generate because of promotions; how this was done previously.

In practice, many retailers have been using a two stage ‘base-times-lift’ approach to generate forecasts for product sales at the SKU level. For example, retailers may generate a baseline forecast using simple exponential smoothing methods and then make adjustments for any incoming promotional event. The adjustments are usually made by brand/category managers. In the literature, a stream of studies have been devoted to help managers improve their adjustment procedure (Fildes et al., 2008; Goodwin, 2002; Lee et al., 2007; Nikolopoulos, 2010). Alternatively, some studies proposed model-based forecasting system estimate the adjustment (Cooper et al., 1999). Other studies including Kuo (2001), Aburto and Weber (2007) and Gur Ali et al. (2009) proposed machine learning algorithms which include the promotional information of the focal product. ([Huang, Fildes et al. 2014](#_ENREF_27)) is the first study which directly incorporates the promotional information of not only the focal product but also of the competitive products within the same product category. ([Ma, Fildes et al. 2016](#_ENREF_39)) further included the promotional information cross categories though using models with different structures and specification strategies.

The issue of structural breaks and forecast bias.

All these studies assume that the effect of the promotional activities does not change over time. In practice, however, this may not be true due to the impact of many influencing factors including the change of economic conditions, new legislation, the change of consumer tastes, and media habits, new competitor entry etc. ([Wildt 1976](#_ENREF_59), [Wildt and Winer 1983](#_ENREF_60)). Under such a circumstance, the model may potentially be subject to structural breaks which is defined as large change in the model with respect to the constant term or/and the parameter coefficients ([Armstrong 2001](#_ENREF_5)). As a result, the model may potentially produce biased and less accurate forecasts. The issue of structural breaks and forecast bias have been intensively address in the economics literature. For example, Pesaram and Timmerman xxx

In this study, we take into account the potential issue of structural breaks and forecast bias using two different approaches: estimation window combining and intercept correction. In the intercept correction approach, we try to identify the existence of structural breaks, estimate the magnitude of forecast bias in the forecasting origin, and then make adjustments to the out-of-sample forecasts. In the estimation window combining approach we generate a set of forecasts produced by the model of the same specification but with different estimation windows. We may expect to obtain more accurate forecast by combining these forecasts.

The rest of the paper is arranged as follows: section 2 summarize the findings of previous studies related to the change of the effect of the promotional activities. Section 3 explains the methodology. Section 4 introduces the data and experimental design. The last sections show the preliminary results.

**Literature review**

2.1 Existing forecasting studies

In practice, many retailers produce forecasts for retailer product sales at the SKU level using a ‘base-times-lift’ approach. In this approach, retailers first generate the ‘baseline’ forecasts using univariate methods with the data when the focal product is not being promoted. They then make adjustments to the baseline forecasts if there is any incoming promotional event in the future (Fildes et al., 2008; Fildes et al., 2009). The univariate models being used are usually simple such as the simple exponential smoothing method, though evidence suggests that the method can be hard to beat for the forecast period when the focal product is not being promoted (Gür Ali et al.. (2009). The adjustments to the incoming promotional event, which are usually done by brand/category managers, are prone to systematic bias (Fildes et al., 2009; Franses and Legerstee, 2010). A stream of studies has been devoted to improve the adjustment by helping managers with their judgmental procedure. (e.g., Trapero Arenas et al., 2013; Fildes et al.). Some other studies try to improve the adjustment with model-based forecasting systems which integrate the information of promotional event conditions and store/category features (Cooper et al. (1999, Cooper and Giuffrida, 2000; Trusov et al., 2006). An intrinsic limitation for this type of methods (i.e., first generate baseline forecast and then make adjustment) is that they produce forecasts separately considering whether the focal product is being promoted or not, and as a result, the information when the focal product is being promoted are naturally overlooked when forecasting the sales of the product when the product is not being promoted, and vice versa.

Some recent studies proposed more sophisticated models to directly forecast the product sales (other than adding the adjustment to the baseline forecast) and also take into account the promotional information of the focal product. Kuo (2001) and Aburto and Weber (2007) evaluated the performance of neural network algorithms in forecasting supermarket food products. [Gür Ali, SayIn et al. (2009)](#_ENREF_22) proposed support vector machine methods and regression tree methods to forecast retailer product sales at SKU level. Divakar et al. (2005) built the CHAN4CAST system which employed models of the regression form with information including past sales, trend, prices and promotions of the focal brand and the major competitors, and seasonality etc. However, the forecasting system were built at the brand/company level and only focused on the interaction between two brands (e.g., Coke and Pepsi). The situation for a retailer becomes very different as there can be hundreds of items competing with each other in a typical product category at any time (Cooper et al. 1999). Huang, Fildes and Soopramanien (2014) is the first study in the forecasting literature which proposed effective forecasting methods for retailer product sales at the SKU level with a formal procedure to select, refine, and incorporate competitive prices and competitive promotions from hundreds of competing items within the same product category. They built generate-to-specific econometric models with the most valuable competitive promotional information either selected by a LASSO/stepwise procedure or condensed by a principle component analysis. The method generated substantially more accurate forecasts across a range of product categories. Ma et al. (2016) further integrated the promotional information not only from the same product category (e.g., intra- category information) but also from other categories (e.g., inter- category information). Their proposed method relies on the LASSO algorithm for both variable selection and model specification.

2.2 The problem of structural break.

Ref which indicates that the effect of promotions may change over time

The effect of promotional activities

A large number of studies have been devoted to exploring the effect of promotional activities (e.g., Blattberg, Briesch et al. 1995, Van Heerde, Gupta et al. 2003). Promotions can significantly increase the short term sales of the focal product (Blattberg, 1995). Promotions have impact on the product sales not only of the focal product but also of complementary and competitive products. (e.g. (Frank and Massy, 1967; Kumar and Leone, 1988; Moriarty, 1985, Mulhern and Leone, 1991; Walters 1988, 1991). More recent studies found that the impact of promotions cross-categories to be asymmetrical as promotions on national brands have much stronger effect on store-label brands (Wedel and Zhang 2004). Promotions also have the dynamic effect. For example, promotions may either accelerate customers’ consumption (Aliwadi and Neslin, 1998) or postpone their purchases if customers can anticipate the promotional events (Van Heerde et al., 2003).

More recent studies focus on the change of the effect of the promotional events.

The effect of promotional activities can change over time

Some studies tend to explore the presumed ‘constant’ effect of the promotional activities on the product sales (or consumer preferences) under specific circumstance (e.g., [Hoch, Kim et al. 1995](#_ENREF_25), [Bijmolt, Heerde et al. 2005](#_ENREF_8)). There are many studies which have been devoted into exploring the changing effects of marketing activities (e.g. [Little 1966](#_ENREF_38), [Morrison 1966](#_ENREF_44), [Myers and Nicosia 1970](#_ENREF_47), [Myers 1971](#_ENREF_46), [Houston and Weiss 1975](#_ENREF_26), [Monroe and Guiltinan 1975](#_ENREF_43), [Moinpour, McCullough et al. 1976](#_ENREF_42), [Wildt 1976](#_ENREF_59), [Wichern and Jones 1977](#_ENREF_58), [Winer 1979](#_ENREF_61), [Mahajan, Bretschneider et al. 1980](#_ENREF_40)). The effect of the marketing activities may change due to exogenous factors, for example, economic condition, legislation, consumer tastes, media habits, competition, and advertising etc. ([Wildt 1976](#_ENREF_59), [Wildt and Winer 1983](#_ENREF_60)). It is generally known that the effects of the marketing mix variables will change with different stages of the product life cycle ([Mahajan, Bretschneider et al. 1980](#_ENREF_40)). For instance, marketing theory suggests that the elasticities for marketing instruments (e.g. advertising, price, service, product quality, and packaging) are the highest at the growth stage of the product and the lowest at the maturity stage of the product ([Kotler 1997](#_ENREF_33)). The introduction of new products (especially the store-owned brand) may decrease the promotional elasticity of the premium national brand and increase the promotional elasticity of the second tier national brand ([Nijs, Dekimpe et al. 2001](#_ENREF_48), [Van Heerde, Srinivasan et al. 2008](#_ENREF_56)).

Intensive promotions can reduce consumers’ reference price ([Lattin and Bucklin 1989](#_ENREF_34), [Lichtenstein and Bearden 1989](#_ENREF_37), [Kalwani, Yim et al. 1990](#_ENREF_30), [Kalwani and Yim 1992](#_ENREF_29), [Foekens, S.H. Leeflang et al. 1999](#_ENREF_20), [Kopalle, Mela et al. 1999](#_ENREF_31), [Levy, Grewal et al. 2004](#_ENREF_36)), which accordingly changes the effects of promotions. For example, consumers may find the promotions less attractive if the products are promoted more frequently than before. The introduction of a new distribution channel can change the market response structure ([Verhoef, Neslin et al. 2007](#_ENREF_57)). For example, consumers may collect information in the newly constructed channel and adjust their reference price accordingly. The introduction of a new loyalty program can change the market response structure ([Leenheer, van Heerde et al. 2007](#_ENREF_35)). For example, retailers may launch promotional events to attract consumers from their competitors. However, when consumers become loyalty program members of a specific retailer, they receive saving rewards and direct discounts, and may find the promotions in other retailers less attractive. Accordingly, the termination of the existing loyalty program also changes the market response structure ([Melnyk and Bijmolt 2007](#_ENREF_41)). The relationship between product sales and the marketing mix variables may also change over time due to the evolving market structure specific to the retailer sales at the UPC level. As was discussed in section 3.2, the UPC set in the product categories may change considerably because ofthe introduction of new product UPCs or termination of existing product UPCs and changes in the assortment policies by the retailer (e.g. the retailer may decide to increase or reduce the number of UPCs in the product category) ([Bell, Bonfrer et al. 2005](#_ENREF_7)). That is, a single product UPC is competing with different sets of competitive products as time goes by, and the effects of the promotions on the focal product may change accordingly.

Foekens, S.H. Leeflang et al. ([1999](#_ENREF_20)) extended the original SCAN\*PRO model to incorporate the time-varying effects of the marketing mix variables. In the extended model, the parameters of the marketing mix variables are functionally related to historical information of the focal brand and other competitive brands. For example, the intercept for the store and the price elasticity of the focal brand are related to previous price discounts of the focal brand and the competitive brands; the elasticities of the non-price promotions for the focal brand are related to the time since the most recent promotion for the focal brand and the competitive brands. The model aims to capture how the effects of the marketing mix variables change over time so that managers can allocate the marketing budget more efficiently. Kopalle, Mela et al. ([1999](#_ENREF_32)) also extended the SCAN\*PRO model in a similar manner to investigate the dynamic impact of promotions on the baseline sales. In their extended SCAN\*PRO model, the effects of price reductions are assumed to change according to previous discounting history. The results show that promotions increase the concurrent product sales but reduce the baseline sales.

However, only a few early studies have attempted to taken into account the change of the effect of marketing activities (e.g. advertising) over time in forecasting product sales ([Mahajan, Bretschneider et al. 1980](#_ENREF_40)). [Cooley and Prescott (1976)](#_ENREF_14) proposed models which allows the parameter to change in an autoregressive manner, say, , , where and are uncorrelated error terms. In an alternative form the parameters were modelled as a function of a constant term with a disturbance term, e.g., , where is the error term. The autoregressive variation model has been applied to capture how the effects of advertising change over time ([Pekelman and Edison 1980](#_ENREF_49)), but the random variation function was rarely used because it does not track the changing effects of the marketing mix variables over time (Wildt and Winer 1983).

A discussion: the effect of promotions change over time due to other influencing factors, but these factors are not included in the model. Or, the bias originate from the fact that there are omitting variables in the model and the impact of these variables change over time.

**Section 3: structural break, forecast bias, and forecast accuracy**

When the effectiveness of the price and promotions on product sales change, conventional econometric models with constant parameters will be subject to structural break which is defined as large changes in the model’s parameters ([Allen and Fildes 2001](#_ENREF_2)). The parameter estimates of the models become the weighted average of the true parameter before and after the break. The forecasts generated by the model will be subject to bias and may not be as accurate as it is supposed to be. The impact of structural break on the model’s forecasting performance has been addressed by many studies in the economics literature (e.g. [Cooper and Nelson 1975](#_ENREF_15), [Muellbauer 1994](#_ENREF_45), [Hendry 1995](#_ENREF_23), [Clements and Hendry 1999](#_ENREF_13), [Pesaran and Timmermann 2007](#_ENREF_50), [Castle, Doornik et al. 2008](#_ENREF_10)).

[Pesaran and Timmermann (2007)](#_ENREF_50) demonstrate the impact of a structural break within the estimation sample on the model’s forecasting performance. Suppose that we have the data from week 1 to week *T,* i.e., and we assume that a structural break occurs at the date of (). For example, the parameter for the price change variable changes from to . In reality, this may be caused by the impact of many factors including a new brand entry, a new advertisement, and the change of the temperature (especially for frozen drinks product) etc. We assume the real demand to be represented as follows:

where, is an indicator which equals to 1 when and 0 otherwise. and are respectively the vectors of the dependent variable and the explanatory variables (e.g., price change) at time *t*. and are the parameter coefficients before and after the structural break, and we assume that . is the error term, and we assume . We also assume that the variance of the error term shifts from to after the time of . We denote that *m* as the first observation in the estimation sample.

If we estimate a model which is congruent with the demand (e.g., ) based on the estimation sample (i.e., from data *m* to *T*), the OLS estimate will be:

where and are respectively the matrices of the explanatory variables and the dependent variable with the observations from *m* to *T*. is not an unbiased estimate of but a weighted average of and . In this case, as we assume that there is no structural break after the time *T[[1]](#footnote-1)*, the true demand will remain as in the out-of-sample data and the out-of-sample forecasting error at the time of *T*+1 is:

The expected value of the error, which is , would not be zero and the forecasting error is not unbiased. Deterministic mean

**Section 3: The method**

In this study, we propose a method of three stages to generate more accurate forecasts by mitigating the issue of structural breaks and forecast bias. We first develop the Autoregressive Distributed Lag (ADL) model, with and without competitive price and promotional information (Huang et al., 2014). To incorporate the competitive price and promotional information, we implement the Least Absolute Shrinkage and Selection Operator (LASSO) following Huang et al. (2014). The LASSO algorithm enable us to select the most relevant competitive promotional information while keep the model in an appropriate size. Alternative, we also implement the principle component strategy proposed by [Stock and Watson (2002)](#_ENREF_53). This strategy pools information across all the competitive explanatory variables and condense them into a small number of diffusion factors.

In the second stage, we incorporate the refined competitive information (e.g., selected explanatory variables or constructed diffusion index) into econometric forecasting models. We develop the Autoregressive Distributed Lag (ADL) model following a general-to-specific modelling strategy ([Hendry 1995](#_ENREF_23)). The ADL model has the advantage of taking into account the carryover effect of the price and promotional variables, and it is transparent with a simple regression style model structure, which benefits the users ([Fader and Hardie 2005](#_ENREF_17)). It also has good interpretability compared to “black box” machine learning approaches which can hardly be understood by brand/category managers. Also the general-to-specific modelling strategy ensures the parsimony and data congruence of the model. Therefore the general-to-specific ADL model is one of the most popular methods in the forecasting literature and it has exhibited superior forecasting performance in other areas including manufacturer sales, tourism, and air passenger flows (see [Albertson and Aylenb 2003](#_ENREF_1), [Song and Witt 2003](#_ENREF_52), [Fildes, Wei et al. 2011](#_ENREF_19)).

In this study, we start with a general model assuming that it properly describes the salient features of the data generating process, and then simplify the general model by seeking out valid parsimonious restrictions. 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 (Huang et al, 2014):

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

is the number of competitive price variables selected by the variable selection methods

is the number of competitive promotional variables selected by the variable selection methods

is the four-week-dummy variable  
 is the dummy variable for the calendar event at week . The dummy variable represents the week of the calendar event when , , and the week before the event if . takes the values from 1 to 9 representing all the calendar events *[[2]](#footnote-2)*

are the parameters  
 is the error term and we assume

is the order of the lags[[3]](#footnote-3).

In the final stage, we take into account of the issue of structural break and forecast bias. These methods are all based on the models which have develop at the first two stage. We first introduce the estimation window combining technique.

When we know the model is subject to structural break, a conventional approach is to estimate the model using the data after the structural break. Suppose we have the DGP as described in section 2.2, i.e,, for the data of . If the time of the structural break is known, the model could simply be estimated based on the data after the break, i.e., [1: *T*] and the model will not be subject to the structural break. The first limitation of this strategy is that we usually do not know the location of the structural break (i.e., ). Statistic tests have been proposed for the purpose (e.g., [Chow 1960](#_ENREF_11), [Andrews 1993](#_ENREF_3), [Andrews and Ploberger 1994](#_ENREF_4), [Bai and Perron 1998](#_ENREF_6)). However these tests may not be reliable because of their limitations e.g., they may assume there is no change in the error variance for the model, or known number of multiple structural breaks before conducting the test etc. In practice, even we know the location of the structural break, we may still need to include the pre-break data because we may not have enough observations to estimate the model if the structural break occurs close to the end of the estimation period.

The first method is that we may take a trade-off between the forecast bias and the forecast error variance. [Pesaran and Timmermann (2007)](#_ENREF_50) proposed to combined the forecasts generated by the same model but estimated with different time windows. Under such a condition, the forecasts generated by some models (e.g., estimated with the data somehow before the structural break) will be biased but with smaller error variance (because more information, e.g., the data before the structural break, has been used in the estimation of the model), and the forecasts generated by other models (e.g., estimated with the data after the structural break) will have inflated forecast error variance (because of the omission of the data before the structural break during the estimation of the model) but with no forecast bias. If we combine the forecasts generated by these models, we may have more accurate forecast results because we may expect a better trade-off between the forecast error variance and the forecast bias which both contribute to the loss function of the retailer. For example, suppose that we have the forecasting error as described in section 3:

Thus the forecasting error measure Mean Square Forecasting Error (MSFE) at the time of *T*+1 conditional on is as follows ([see equation (7) in Pesaran and Timmermann 2007](#_ENREF_50)):

where , and is a diagonal matrix where the first diagonal places are and the remaining diagonal places are . The can be decomposed as ([see equation 8 in Pesaran and Timmermann 2007](#_ENREF_50)):

where

In this equation, is the squared forecast bias, and is the efficiency term ( is the forecasting error variance). When one additional observation is added in the estimation sample, the change in the error measure becomes:

where is the MSFE calculated with an estimation window with one extra observation compared to . In this equation, the term () is always larger than or equal to zero, and the sign of depends on the sign of (i.e. ). For example:

where and

Thus the sign of depends on the sign of which is the proportion of the change in the error variance compared to the variance after the structural break (i.e. ). For example, if , , and will be smaller than or equal to ; if , , and might be either smaller or larger than .   
   
Therefore, when we include additional observations before the structural break in the model estimation, we will have increased forecast bias, which is the very problem of structural break. However, we will have the forecasting error variance either increased (if ) or decreased (if ) depending on whether the variance of the error terms in the DGP decrease or increase after the structural break. Therefore, we can either see an increased or decreasd when we include more data before the structural break depending on a trade-off between the rise in the squared forecast bias (i.e.) and the potential fall in the efficiency term (i.e. , if ).

Then we can combine the forecasts generated by the models estimated with different time windows. Specifically, If we denote the whole estimation period as , we can estimate the model using the latest observations (i.e. the data in ) to generate the first set of *h*-step-ahead forecasts as:

We can then estimate the model with the latest observations (i.e. ) and generate the second set of the *h*-step-ahead forecasts:

We can repeat this process by adding more observations until we use all the observations in the estimation sample (i.e. ) to generate the *h*-step-ahead forecast:

Eventually, we would have calculated the final *h*-step-ahead forecasts by taking an average of the () sets of *h*-step-ahead forecasts based on, for example, an equal weighting scheme:

In the combination, can be arbitrarily chosen as long as we can ensure there are enough observations to estimate the model and there are enough variations in all the explanatory variables. Pesaran, Schuermann et al. ([2009](#_ENREF_51)) found that this approach improved the forecasting performance for the random walk with a drift model and the VAR model which are both subject to multiple structural breaks. In this study, we apply the estimation window combining approach with equal weights for the results obtained from each estimation window because it usually generates better performance compared to alternative combining schemes and easy to implement (Stock and Watson, 2001).

Alternatively, we may apply the intercept correction (IC) method. This method firstly identifies the existence of the structural break and then estimates the magnitude of the consequent forecast bias at the forecast origin. It then offsets the forecast bias by specifying non-zero values for the model’s errors in the forecast period. That is, it adds the estimated forecast bias back to the out-of-sample forecasts. The intercept correction technique may potentially improve the forecasting accuracy by mitigating the forecast bias. The technique has been applied in making adjustments for macro-economic forecasts ([Clements and Hendry 1994](#_ENREF_12)).

Once we have identified that the model is subject to structural break, we assume that the model is subject to forecast bias. The estimate of the bias could be done following different schemes. For example, we may estimate the forecasting bias as the predictive error at the forecast origin (i.e., , where *T* is the last observation in the estimation window). Alternatively, we may estimate the bias as the average value of an ad hoc number of predictive errors before the forecast origin. (e.g. , where *i* is arbitrarily chosen). The approach will then add the estimated bias back to the out-of-sample forecasts following various correction strategies. Clements and Hendry ([1999](#_ENREF_13)) demonstrated the analytical characteristics of various correction strategies using an example of VAR(1) model with a time trend, i.e., . Suppose that the model is subject to structural break and the forecast bias is estimated as . Denote , , and , as the corrected *h*-step-ahead forecast by the intercept correction technique following various strategies. The intercept correction approach could first makes 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 already adjusted, and so forth. The adjusted *h*-step-ahead forecast is described as . This equation can be re-written recursively as , where is the original *h*-step-ahead forecast. An alternative strategy is to only adjust the one-step-ahead forecast, and . This equation can be re-written recursively as . Another correction strategy makes adjustments to the *h*-step-ahead forecast using the full amount of the forecast bias. That is, .

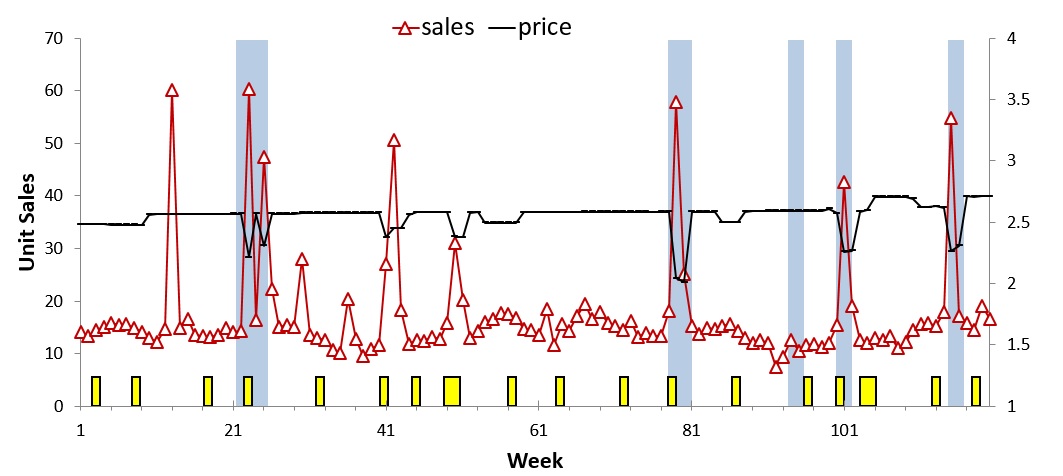
The intercept correction technique can potentially reduce the forecasting bias if the forecasting bias is estimated properly. Further, whether it could improve the forecasting accuracy is an empirical question because this approach comes with the cost of inflated forecasting error variance. Clements and Hendry ([1999](#_ENREF_13)) derived analytically the forecast bias and the inflated forecasting error variance for the VAR(1) model described the various correction strategies. Their findings show that these correction strategies have their own advantages and limitations in terms of the reduced forecasting bias and inflated forecasting error variance depending on the details of the structural break such as which of parameters have changed and whether there are multiple breaks etc.

In this study, we estimate the forecast bias as the (equally weighted) average value of four predictive errors before the forecast origin, and we make adjustments to the *h*-step-ahead forecast using the full amount of the forecast bias. We choose to implement the intercept correction approach both indiscriminately and discriminately depending on the test for structural break. In the discriminated intercept correction strategy, we first conduct the Chow ([1960](#_ENREF_11)) test sequentially to investigate whether the model is subject to structural break based on most of the observations in the estimation sample (i.e. 90% central observations). If the test is rejected for any of the observations, the model is identified as being subject to structural break. A very small *p*-value (i.e. 0.005) is used to mitigate the multiple testing problem in detecting the existence of the structural break. The intercept correction technique will be implemented if and only if the model is identified as being subject to structural break.

**Section 4: The data**

In this study, we evaluate our model using the IRI dataset for which a descriptive article can be found in [Forni and Reichlin (1996)](#_ENREF_21)[[4]](#footnote-4). The IRI dataset contains weekly data at the SKU level including unit sales, price, features and displays. We select 224 SKUs in 29 product categories in one large store. The SKUs we include in our experiment all have positive movement for at least 90% of time. Figure 1 shows the sales graph for a typical product in the Juice category.

Figure 1. The sales, price, and promotional information for SKU in the Juice category.



**Section 5: The candidate models**

In practice, many retailers tend to use the base-times-lift approach where the forecasts are generated using simple univariate method (e.g., simple exponential smoothing) with the adjustment by category managers for incoming promotional events. This approach has been described in [Gür Ali, SayIn et al. (2009)](#_ENREF_22) as:

Where represents the baseline forecast at week generated by the simple exponential smoothing model. represents the actual sales at the previous week when the focal product was not promoted. The parameter is estimated by minimizing the mean squared error in the estimation period. The adjustment is estimated as the increased sales by the most recent promotion of the focal product.

We include the base-times-lift approach as one of the benchmark models. We also include the variants of the ADL model with and without competitive price/promotional information, as proposed by Huang et al. (2014). e.g., the ADL-own model which only contains the price/promotion information of the focal product, the ADL model which incorporates the competitive price/promotion selected by the LASSO algorithm, and the ADL-Diffusion Index (ADL-DI) model which incorporates the competitive price/promotion information via principle component analysis.

Our proposed candidate models include the variants of the ADL models with the techniques of both intercept correction and estimation-window combing. These two techniques are implemented in a selective manner considering the existence of structural breaks. For example, we first conduct the Chow test sequentially for each observation in the estimation period to investigate if the model is subject to structural break. If the null hypothesis of no structural break is rejected for any one of the observations, the model is considered as being subject to structural break. A very small *p*-value (i.e. 0.005) is used for the Chow test to mitigate the multiple testing problem in detecting the structural break. The techniques of intercept correction and estimation-window combing will only be implemented when the model is subject to structural break. The benchmarks and the candidate models are listed in Table 1.

Table 1 shows the candidate models:

|  |  |
| --- | --- |
| Base-times-lift | Industrial practice, Simple-exponential smoothing with adjustments based on the effect of the most recent promotional event |
| ADL-own | ADL model, with the promotional variables of the focal product only |
| ADL | ADL model, based on the variables retained by LASSO |
| ADL-DI | ADL model, based on the diffusion factors constructed by principle component analysis |
| ADL-EWC | ADL model with estimation window combining |
| ADL-DI-EWC | ADL-DI model, with estimation window combining |
| ADL-IC | ADL model with intercept correction |
| ADL-DI-IC | ADL-DI model, with intercept correction |
|  |  |

**Section 6: The experimental design**

In this study, we evaluate the models with 30 rolling forecast origins with multiple forecast horizons. This ensures the results to be more robust to randomness and systematic business cycle effects ([Fildes 1992](#_ENREF_18), [Stock and Watson 2002](#_ENREF_54), [Stock and Watson 2002](#_ENREF_53)). We estimate the models with a moving window of 120 weeks and forecast one to weeks ahead, where is 1, 4, and 12 as we consider typical ordering and planning periods. We then re-estimate the model with updated data by including the data in the latest week and dropping the data in the earliest week. We repeat this process until we have used all the data in the remaining estimation sample. Therefore, in the experiment, we have 30 sets of one to weeks ahead forecast in total. When the lead times are greater than one, we use the actual value of the explanatory variables (e.g., price and promotion etc.) and the forecasted values of the lagged dependent variables. In practice, promotional variables are usually known to retailers as they are included as one part of the agreed promotional plan between retailers and manufacturers. We specify the ADL models based on the data from week 1 to week 150 to represent the model which would ideally be specified with the foreknowledge of the data ([Fildes, Wei et al. 2011](#_ENREF_19)). An alternative way to evaluate the models is to re-specify the model for each rolling event based on each moving estimation window (Ma et al, 2016).

We follow Huang et al. (2004) to evaluate the models’ forecasting performance using three error measures: the Mean Absolute Scaled Error (MASE), the MAPE, the symmetric Mean Absolute Percentage Error (sMAPE), In this study, the MAPE and the symmetric MAPE for data series *s* with forecast horizon for the rolling event are shown as follows:

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[[5]](#footnote-5).

The MASE represents the “weighted” arithmetic mean of the MAE compared to the variations in the estimation sample ([Hyndman and Koehler 2006](#_ENREF_28)). This is calculated across data series with forecast horizon for the rolling event is as follows:

Within the equation for , the numerator, i.e., , is the MAE for data series with forecast horizon for the rolling event, while the denominator is the sum of the one-step-ahead predicted errors by the no-change naïve model in the estimation sample. is the actual value for data series in the estimation period for the rolling event, and is the total number of observations in the estimation period.

The three error measures are all approximations of the unknown loss function of the retailer, and they penalize the forecast errors with different aspects. 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).

**Section 7: Results**

We first examine the overall forecasting performance of the models across all the SKUs for various error measures. Table 2 shows the results for three error measures across all the 224 products for the average forecasting horizon of 1-12 weeks. Table 2 has the following indications: 1) the Base-times-lift benchmark has been outperformed by all the candidate models regardless of the error measure; 2) the ADL-own model, which incorporated the promotional information of the focal product, is outperformed by the ADL model and the ADL-DI model, which incorporated the promotional information not only from the focal product but also from other competitive products in the same product category. This finding is consistent with Huang et al (2014) and Ma et al (2016); 3) the ADL-EWC model and the ADL-IC model outperform the ADL model, and the ADL-DI-EWC model and the ADL-DI-IC model outperform the ADL-DI model. Thus the ADL model and the ADL-DI model, which both incorporate competitive promotional information within the same product category, can be improved by using the estimation combining approach and the intercept correction method; 4) the ADL-own-EWC model and the ADL-own-IC model outperform the ADL-own model. Therefore, even when competitive promotional information is not available, we can still improve the forecasting performance of the ADL-own model with the estimation combining approach and the intercept correction method. That is, the new proposed model can also benefit manufacturers whom we do not assume to have the price and promotional information of their competitors.

Table 2 Forecast performance of the various models based on the 1-12 weeks forecast horizon.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Candidate models | MAPE | RANK | MASE | RANK | SMAPE | RANK |
| Base-times-lift | 54.1% | 10 | 0.820 | 10 | 41.7% | 10 |
| ADL-own | 48.8% | 9 | 0.740 | 9 | 35.0% | 9 |
| ADL | 45.8% | 6 | 0.713 | 5 | 34.0% | 7 |
| ADL-own-EWC | 47.9% | 8 | 0.734 | 8 | 34.8% | 8 |
| ADL-own-IC | 47.4% | 7 | 0.719 | 7 | 34.0% | 6 |
| ADL-IC | 44.7% | 3 | 0.703 | 1 | 33.3% | 1 |
| ADL-EWC | 45.1% | 5 | 0.712 | 3 | 33.8% | 3 |
| ADL-DI | 44.8% | 4 | 0.714 | 6 | 33.9% | 4 |
| ADL-DI-IC | 44.3% | 2 | 0.710 | 2 | 33.7% | 2 |
| ADL-DI-EWC | 44.1% | 1 | 0.712 | 4 | 34.0% | 5 |

**Forecasting performance comparison when the focal product is and is not being promoted.**

**Forecasting performance regarding forecasting horizons of 1 week and 1-4 week.**

**Section 8: Conclusion and Future research**

Grocery retailers needs accurate sales forecasts to improve their inventory management performance. In practice, retailers are facing intense competitions and spending heavily on price reductions and promotional activities, which has substantially increased the variation in the product sales. Previous studies proposed to incorporate the price and promotional information, not only from the focal product but also from other competitive products, in forecasting retailer product sales. These studies assumed the effectiveness of price and promotions to be constant. However, in practice, the effectiveness of price reductions and promotions may change due to unobserved influencing factors including economic conditions, the entry of new brands, competition, and the change of consumers’ tastes etc. As a result, the models may potentially generate biased forecasts due to structural breaks.

In this study, we propose a three-stage method to forecast retailer product sales at the SKU/store level. We take into account the potential issue of forecast bias by using recently developed techniques including the estimation window combining strategy and the intercept correction approach. Our results show that we can improve the forecasting accuracy of the econometric models by using these methods regardless of whether competitive promotional information have been incorporated.

There are alternative method which also mitigate the problem of structural break and forecast bias. [Castle, Doornik et al. (2008)](#_ENREF_10) and [Hendry and Krolzig (2001)](#_ENREF_24) proposed the saturation approach where the regression model was initially incorporated dummy variables for each observation and then recursively reduced by an algorithm called *Autometrics* based on the General-to-specific modelling strategy. The ultimate model will not be subject to structural break and thus would be expected to product unbiased forecast. However the method comes with the cost of losing information (e.g. the observations offset by the retained dummy variables) and its performance becomes an empirical question, and we leave this to our next research question.

In this study, we find the estimation window combining method and the intercept correction method can improve the models’ forecasting performance regardless of whether the competitive promotional information has been incorporated. Ma et al. (2016) proposed models which further integrate both the intra and the inter category promotional information. Thus it is promising to implement the estimation window combining method and the intercept correction method to the models with the intra and the inter category promotional information. However, the model in Ma et al. (2016) consists of three stages where each sequent stage bases on the error of the previous stage and it is not straightforward to apply the methods in this study to their model.

Reference:

Albertson, K. and J. Aylenb (2003). "Forecasting the behaviour of manufacturing inventory " International Journal of Forecasting **19**(2): 299-311.

Allen, P. G. and R. Fildes (2001). Econometric forecasting. Principles of Forecasting: A Handbook for Researchers and Practitioners. J. S. Armstrong. Boston, Kluwer Academic Publishers.

Andrews, D. W. K. (1993). "Tests for Parameter Instability and Structural Change with Unknown Change Point." Econometrica **61**: 825-851.

Andrews, D. W. K. and W. Ploberger (1994). "Optimal tests when a nuisance parameter is present only under the alternative." Econometrica **62**: 1383-1414.

Armstrong, J. S. (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners, Kluwer Academic Publishers.

Bai, J. and P. Perron (1998). "Estimating and Testing Linear Models with Multiple Structural Changes." Econometrica **66**: 47- 78.

Bell, D. R., et al. (2005). "Recovering Stockkeeping-Unit-Level Preferences and Response Sensitivities from Market Share Models Estimated on Item Aggregates." Journal of marketing research **XLII**(May).

Bijmolt, T. H. A., et al. (2005). "New Empirical Generalizations on the Determinants of Price Elasticity." Journal of marketing research **XLII**: 141-156.

Blattberg, R. C., et al. (1995). "How promotions work?" Marketing Science **14**(3).

Castle, J. L., et al. (2008). "Model Selection when there are Multiple Breaks." Working paper No. 407, Economics Department, University of Oxford.

Chow, G. C. (1960). "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." Econometrica **28**(3).

Clements, M. B. and D. F. Hendry (1994). Towards a theory of economic forecasting. Nonstationary Time Series Analysis and Cointegration. C. P. Hargreaves, Oxford University Press.

Clements, M. P. and D. F. Hendry (1999). Forecasting non-stationary economic time series. London, The MIT Press.

Cooley, T. F. and E. C. Prescott (1976). "Estimation in the Presence of Stochastic Parameter Variation." Econometrica **44**(1): 167-184.

Cooper, J. P. and C. R. Nelson (1975). "The Ex Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models and Some Results on Composite Predictors." Journal of Money, Credit and Banking **7**(1).

Corsten, D. and T. Gruen (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).

Fader, P. S. and B. S. Hardie (2005). "The value of simple models in new product forecasting and customer-base analysis." Applied Stochastic Models in Business and Industry **21**.

Fildes, R. (1992). "The evaluation of extrapolative forecasting methods." International Journal of Forecasting **8**: 81-98.

Fildes, R., et al. (2011). "Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures." International Journal of Forecasting **27**(3): 902-922.

Foekens, E. W., et al. (1999). "Varying parameter models to accommodate dynamic promotion effects." Journal of econometrics **89**(1-2): 249-268.

Forni, M. and L. Reichlin (1996). "Lets Get Real: A Dynamic Factor Analytical Approach to Disaggregated Business Cycle." Review of Economic Studies **65**.

Gür Ali, Ö., et al. (2009). "SKU demand forecasting in the presence of promotions." Expert Systems with Applications **36**(10).

Hendry, D. F. (1995). Dynamic Econometrics: Advanced Texts in Econometrics. Oxford, UK, Oxford University Press.

Hendry, D. F. and H.-M. Krolzig (2001). Automatic Econometric Model Selection using PcGets. London, Timberlake Consultants Press.

Hoch, S. J., et al. (1995). "Determinants of store-level price elasticity." Journal of marketing research **XXXII**: 17-29.

Houston, F. S. and D. L. Weiss (1975). "CUMULATIVE ADVERTISING EFFECTS: THE ROLE OF SERIAL CORRELATION\*." Decision Sciences **6**(3): 471-481.

Huang, T., et al. (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. and A. B. Koehler (2006). "Another look at measures of forecast accuracy." International Journal of Forecasting **22**: 679-688.

Kalwani, M. U. and C. K. Yim (1992). "Consumer Price and Promotion Expectations: an Experimental Study." Journal of marketing research **29**(February): 90-100.

Kalwani, M. U., et al. (1990). "A Price Expectations Model of Customer Brand Choice." Journal of marketing research **27**(3): 251-262.

Kopalle, P. K., et al. (1999). "The Dynamic Effect of Discounting on Sales: Empirical Analysis and Normative Pricing Implications." Marketing Science **18**(3): 317-332.

Kopalle, P. K., et al. (1999). "The Dynamic Effect of Discounting on Sales: Empirical Analysis and Normative Pricing Implications." Marketing Science **18**(3): 317.

Kotler, P. (1997). Marketing management, analysis, planning, implementation and control, Prentice Hall.

Lattin, J. M. and R. E. Bucklin (1989). "Reference effects of price and promotion on brand choice behavior." Journal of marketing research **26**: 299-310.

Leenheer, J., et al. (2007). "Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members." international Journal of research in marketing **24**(1): 31-47.

Levy, M., et al. (2004). "Emerging trends in retail pricing practice: implications for research." Journal of Retailing **80**.

Lichtenstein, D. R. and W. O. Bearden (1989). "Contextual Influences on Perceptions of Merchant-Supplied Reference Prices." Journal of Consumer Research: An Interdisciplinary Quarterly **16**(1): 55-66.

Little, J. D. C. (1966). "A Model of Adaptive Control of Promotional Spending." Operations research **14**(6).

Ma, S., et al. (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.

In marketing analytics applications in OR, the modeler often faces the problem of selecting key variables from a large number of possibilities. For example, SKU level retail store sales are affected by inter and intra category effects which potentially need to be considered when deciding on promotional strategy and producing operational forecasts. But no research has yet put this well accepted concept into forecasting practice: an obvious obstacle is the ultra-high dimensionality of the variable space. This paper develops a four steps methodological framework to overcome the problem. It is illustrated by investigating the value of both intra- and inter-category SKU level promotional information in improving forecast accuracy. The method consists of the identification of potentially influential categories, the building of the explanatory variable space, variable selection and model estimation by a multistage LASSO regression, and the use of a rolling scheme to generate forecasts. The success of this new method for dealing with high dimensionality is demonstrated by improvements in forecasting accuracy compared to alternative methods of simplifying the variable space. The empirical results show that models integrating more information perform significantly better than the baseline model when using the proposed methodology framework. In general, we can improve the forecasting accuracy by 12.6 percent over the model using only the SKU's own predictors. But of the improvements achieved, 95 percent of it comes from the intra-category information, and only 5 percent from the inter-category information. The substantive marketing results also have implications for promotional category management.

Mahajan, V., et al. (1980). "Feedback Approaches to Modeling Structural Shifts in Market Response." Journal of marketing **44**: 71-80.

Melnyk, V. and T. H. A. Bijmolt (2007). "Effects of having and terminating a loyalty program: the role of monetary and non-monetary rewards." Working paper New Zealand: University of Waikato.

Moinpour, R., et al. (1976). "Time Changes in Perception: A Longitudinal Application of Multidimensional Scaling." Journal of marketing research **13**(3): 245-253.

Monroe, K. B. and J. P. Guiltinan (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**.

Muellbauer, J. (1994). "The Assessment: Consumer Expenditure." Oxford Review of Economic Policy **10**(2): 1-41.

Myers, J. G. (1971). "The Sensitivity of Dynamic Time-Path Typologies." Journal of marketing research **8**(4): 472-479.

Myers, J. G. and F. M. Nicosia (1970). "Time-Path Types: From Static to Dynamic Typologies." Management Science **16**(10): B584-B596.

Nijs, V. R., et al. (2001). "The Category-Demand Effects of Price Promotions." Marketing Science **20**(1): 1-22.

Pekelman, D. and T. Edison (1980). "Experimentation and Budgeting in Advertising: An Adaptive Control Approach." Operations research **28**(2): 321-347.

Pesaran, H. M. and A. Timmermann (2007). "Selection of estimation window in the presence of breaks." Journal of econometrics **137**: 134-161.

Pesaran, M. H., et al. (2009). "Forecasting Economic and Financial Variables with Global VARs." International Journal of Forecasting **25**: 642-675.

Song, H. and S. F. Witt (2003). "Tourism Forecasting: The General-to-Specific Approach." Journal of Travel Research **42**: 65-74.

Stock, J. H. and M. W. Watson (2002). "Forecasting Using Principal Components from a Large Number of Predictors." Journal of the american statistical association **97**(460): 1167-1179.

Stock, J. H. and M. W. Watson (2002). "Macroeconomic Forecasting Using Diffusion Indexes." Journal of Business & Economic Statistics **20**(2): 147-162.

Van Heerde, H. J., et al. (2003). "Is 75% of the Sales Promotion Bump Due to Brand Switching? No, Only 33% Is." Journal of marketing research **XL**: 481-491.

Van Heerde, H. J., et al. (2008). "Decomposing the Demand for a Pioneering Innovation." Working paer, University of Waikato, Department of Marketing.

Verhoef, P. C., et al. (2007). "Multichannel customer management: Understanding the research-shopper phenomenon." international Journal of research in marketing **24**(2): 129-148.

Wichern, D. W. and R. H. Jones (1977). "Assessing the Impact of Market Disturbances Using Intervention Analysis." Management Science **24**(3): 329-337.

Wildt, A. R. (1976). The empirical investigation of time dependent parameter variation in marketing models. American Marketing Association. E. proceedings**:** 466-472.

Wildt, A. R. and R. S. Winer (1983). "Modeling and Estimation in Changing Market Environments." The Journal of Business **56**(3).

Winer, R. S. (1979). "An Analysis of the Time-varying Effects of Advertising: The Case of Lydia Pinkham." The Journal of Business **52**(4).

1. Clements and Hendry (1999) showed analytically the impact of out-of-sample structural breaks on a VAR model’s forecasting performance. [↑](#footnote-ref-1)
2. 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-2)
3. 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-3)
4. All estimates and analyses in this paper based on Information Resources, Inc. data are by the author and not by Information Resources, Inc. [↑](#footnote-ref-4)
5. 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-5)