**Forecasting Retailer Product Sales at the SKU level with the impact by the unobserved factors**

Tao Huang, Robert Fildes, Didier Soopramanien

Abstract

Retailers need accurate sales forecasts for their inventory management. In this study we propose more effective methods to generate more accurate forecasts by taking into account the unobserved change in the effectiveness of the price and promotional activities. We implement the intercept correction method and the estimation window combining method to mitigate the issue of structural breaks and forecast bias. We evaluate our models for retailer products at the SKU level and we found our proposed new models with intercept corrections with the best forecasting performance.

Key words:

Sales Forecasting, Marketing analytics, Promotion

1. **Introduction**

Retailers have been struggling with the situations of out-of-stocks and overstocks for years. When a product item is out-of-stock, retailers not only immediately lose profits but also may lose the customers in the long term. Customers were believed to purchase alternative products or postpone their purchases when they find the product is out of stock, but studies show that they are more likely to switch to other stores and never come back ([Corsten and Gruen 2003](#_ENREF_19)). In practice, retailers may deliberately increase their inventory level (i.e. to over-stock) to avoid the out-of-stock condition. However, this may significantly raises inventory costs and reduces profits ([Cooper, Baron et al. 1999](#_ENREF_17)). It is estimated that retailers in North American lost $634.1 billion due to out-of-stocks and $471.9 billion due to overstocks just in the year of 2014 ([OrderDynamics 2015](#_ENREF_59)). Under such 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 at the SKU level.

In practice, many retailers forecast product sales using a two-stage ‘base-lift’ approach at the SKU level. Specifically, they produce the baseline forecast using simple exponential smoothing methods and then make adjustments to the baseline forecast for any promotional event over the forecast period. The adjustments are usually made by brand/category managers with their experience. In the forecasting literature, a stream of studies has been devoted to helping managers improve their adjustment procedure (e.g., [Goodwin 2002](#_ENREF_29), [Fildes, Nikolopoulos et al. 2008](#_ENREF_25), [Nikolopoulos 2010](#_ENREF_58)). [Cooper, Baron et al. (1999)](#_ENREF_17) suggested estimate the adjustment with the model-based method and developed the PromoCast® forecasting system. Other studies proposed holistic sophisticated models (e.g., neural networks and machine learning algorithms) which directly generate the final forecasts for product sales (e.g., [Aburto and Weber 2007](#_ENREF_1), [Gür Ali, SayIn et al. 2009](#_ENREF_30)). [Huang, Fildes et al. (2014)](#_ENREF_34) proposed general-to-specific Autoregressive Distributed Lag models which incorporate 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_47)) further integrated the promotional information from the products across different categories.

All the studies mentioned above assume invariant effectiveness of the promotional activities. In practice, this may not be true because of the impact of many influencing factors including the change of economic conditions, the change of consumer tastes, and media habits, new competitor entry etc. ([Wildt 1976](#_ENREF_71), [Wildt and Winer 1983](#_ENREF_72)). The business environment changes quickly and magnificently. A recent example is that the low-price German retailer Aldi which had huge expansions across Europe has opened more than 400 stores in the United States just in the year of 2014 ([Loeb 2015](#_ENREF_46)). Under such circumstance, conventional models may potentially be subject to structural break which is defined as large change in the model with respect to the parameter coefficients ([Armstrong 2001](#_ENREF_8)). As a result, the model may produce biased and less accurate forecasts. The issue of structural break and the opportunity to improve forecasting performance by mitigating the consequent forecast bias have been intensively addressed in the macroeconomics literature ([see Clements and Hendry 1994](#_ENREF_14)). However, this issue has been totally overlooked in the area of forecasting retailer product sales.

In this study, we aim to improve the forecasting performance for retailer product sales at the SKU level by taking into account the unobserved change of the effectiveness of the price and promotional activities. We propose models which generate more accurate forecasts by mitigating forecast bias caused by potential structural break using two different methods. For the Intercept Correction (IC) method, we make adjustments to the out-of-sample forecasts based on an estimate of forecast bias. For the estimation window combining (EWC) method, we take a trade-off between the forecast bias and the forecast error variance by combing the forecasts generated by the same model but with different estimation windows. These two methods have been successfully applied with VAR models in forecasting macroeconomic data. However, the retailer product sales data at the SKU level contain noisy variations caused by promotions which are very different from macroeconomic data in terms of data characteristics. This raises two risks: first, the impact by the forecast bias to the final forecasting accuracy may be submerged in the variation of the sales. Second, the improvement by using these two methods may to be small compared to the variation of the product sales. Therefore, whether we can generate more accurate forecasts for retailer product sales by taking into account the issue of structural break with these methods is an empirical question. Methodologically our study conducts a novel evaluation of the method which leads the combine of dynamic models and variable selection process a dimension forward by taking into account the unobserved change of the effectiveness of the explanatory variables.

We implement the two methods to the two two-stage models in [Huang, Fildes et al. (2014)](#_ENREF_34). e.g., the general-to-specific Autoregressive Distributed Lag (ADL) model with LASSO selection and the ADL model with Diffusion factors constructed by Principle Component Analysis. The results indicate that the IC method can improvement the forecasting performance of the ADL model for almost all scenarios, while the EWC method only moderately improves the forecasting performance of the ADL model for some occasions. Overall, the ADL-IC model generate the most accurate forecasts for retailer product sales at the SKU level. We also see that the ADL-own-IC model with the best forecasting performance when competitive information is not an option. Therefore, our models have practical significance in that they equip retailers as well as manufacturers the tools to forecast product sales and more effectively manage their inventory planning.

The remainder of the paper is arranged as follows: section 2 summarizes previous research findings. Section 3 introduces the issue of structural break and resulted forecast bias when conventional models overlook the change in the effectiveness of the price and promotional activities. In section 4, we introduce two methods from the macroeconomic literature to mitigate the issue of structural break. Section 5 and section 6 describe the data and the candidate models respectively. Section 7 introduces the experimental design for the evaluation. In section 8, we summarize and discuss the evaluation results. In section 9, we draw conclusions and make recommendations for both manufacturers and retailers, and we also address some research limitations and possibilities for future research.

**2. Literature review**

2.1 Forecasting retailer product sales at the SKU level

In practice, many retailers produce forecasts for retailer product sales at the SKU level using the ‘base-lift’ approach. For example, they first generate the ‘baseline’ forecasts using univariate methods but excluding the data when the focal product is being promoted. They then make adjustments to the baseline forecast if there is an incoming promotional event in the future ([Fildes, Nikolopoulos et al. 2008](#_ENREF_25), [Fildes, Goodwin et al. 2009](#_ENREF_24)). The univariate models for the ‘baseline’ forecast are usually simple such as the simple exponential smoothing method, though evidence suggests that these simple models can be hard to defeat when the sales data are relatively stable (e.g., over the forecast period when the focal product is not being promoted) ([Gür Ali, SayIn et al. 2009](#_ENREF_30)). The adjustments to the incoming promotional event, which are usually done by brand/category managers, are prone to systematic bias and associated with high costs ([Fildes, Goodwin et al. 2009](#_ENREF_24)). A stream of studies has been devoted to helping managers with their judgmental procedure ([Arenas, Pedregal et al. 2013](#_ENREF_7)). Some other studies try to improve the adjustment with model-based forecasting systems which estimate the ‘lift’ effect by the promotional event using historical information related to previous promotions, store/category features, and manufacturers ([Cooper, Baron et al. 1999](#_ENREF_17), [Cooper and Giuffrida 2000](#_ENREF_18), [Trusov, Bodapati et al. 2006](#_ENREF_66)). However, there is an intrinsic limitation for these methods of two stages (i.e., baseline with adjustments) because they produce forecasts separately depending on whether the focal product is being promoted or not. Under such circumstance, the information when the focal product is being promoted are inevitably overlooked when forecasting the sales of the product when the product is not being promoted, and vice versa.

Some studies proposed holistic methods which relies on sophisticated data mining and machine learning algorithms. [Aburto and Weber (2007)](#_ENREF_1) evaluated the performance of neural network algorithms in forecasting supermarket food sales. [Gür Ali, SayIn et al. (2009)](#_ENREF_30) proposed support vector machine models and the regression tree models to forecast retailer product sales at the SKU level. They also constructed a range of variables based on the promotional information of the focal product (i.e. average, sum, trend, standard deviation of previous sales and price etc.). Other studies proposed econometric forecasting models which not only benefit from high forecasting accuracy but also interpretability which is critically important for practical application. Divakar et al. (2005) proposed the CHAN4CAST system with dynamic regression models to forecast brand sales for manufacturers/channels. More recent studies tried to increase forecasting accuracy through the direction of incorporating more information. For example, [Huang, Fildes et al. (2014)](#_ENREF_34) included competitive promotional information within the same product category to forecast the sales of the focal product. They dealt with the high dimensionality problem specific to the retailer context at the SKU level with variable selection methods and principle component analysis. [Ma, Fildes et al. (2016)](#_ENREF_47) further integrated the promotional information not only from the same product category of the focal product but also from other related categories. They rely on Granger causality test to find out the relevant product categories and then adopted the LASSO algorithm not only for variable selection but also as a model specification strategy.

2.2 The changing effectiveness of marketing activities

Many studies in the marketing literature have explored the effect of promotional activities. For example, early studies found that promotions significantly increase short-term sales of the focal product (Blattberg, 1995). Evidence also show that promotions have positive (negative) impact on complementary (competitive) products ([Wittink, Addona et al. 1988](#_ENREF_74), [Dekimpe, Hanssens et al. 1999](#_ENREF_21), [Andrews, Currim et al. 2008](#_ENREF_6)). The impact of promotions can be asymmetrical as promotions on national brands have much stronger effect on store-label brands (Wedel and Zhang 2004). Promotions also have dynamic effects. For example, promotions may either accelerate customers’ consumption or postpone their purchases if customers anticipate future promotional events ([Van Heerde, Gupta et al. 2003](#_ENREF_67), [Mace and Neslin 2004](#_ENREF_48)).

Some other studies focused on investigating the change in the effectiveness of the promotional activities (e.g. [Little 1966](#_ENREF_45), [Morrison 1966](#_ENREF_53), [Myers and Nicosia 1970](#_ENREF_56), [Myers 1971](#_ENREF_55), [Houston and Weiss 1975](#_ENREF_33), [Monroe and Guiltinan 1975](#_ENREF_52), [Moinpour, McCullough et al. 1976](#_ENREF_51), [Wildt 1976](#_ENREF_71), [Wichern and Jones 1977](#_ENREF_70), [Winer 1979](#_ENREF_73), [Mahajan, Bretschneider et al. 1980](#_ENREF_49)). They argue that the effectiveness of promotions may change because of various reasons. For example, the change in economic condition, legislation, consumer tastes, media habits, and advertising etc. ([Wildt 1976](#_ENREF_71), [Wildt and Winer 1983](#_ENREF_72)). Also, the effectiveness of promotions may change with the different stages of the product life cycle ([Mahajan, Bretschneider et al. 1980](#_ENREF_49)). 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_40)). The impact by other influencing factors such as new competition may also change the effectiveness of some marketing activities. e.g., the introduction of new products (especially the store-owned brand) decrease promotional elasticities of premium national brands and increase promotional elasticities of the second tier national brands ([Nijs, Dekimpe et al. 2001](#_ENREF_57), [Van Heerde, Srinivasan et al. 2008](#_ENREF_68)). Further evidence indicates that intensive promotions can make consumers less responsive to price promotions by reducing consumers’ reference price ([Lattin and Bucklin 1989](#_ENREF_41), [Lichtenstein and Bearden 1989](#_ENREF_44), [Kalwani, Yim et al. 1990](#_ENREF_37), [Kalwani and Yim 1992](#_ENREF_36), [Foekens, Leeflang et al. 1999](#_ENREF_28), [Kopalle, Mela et al. 1999](#_ENREF_39), [Levy, Grewal et al. 2004](#_ENREF_43)). The introduction of a new distribution channel (e.g., online website) can also reduce the effectiveness of price promotions on the original channel ([Verhoef, Neslin et al. 2007](#_ENREF_69)). This is because that consumers may get more easily to collect information through the newly constructed channel and then reduce their reference price. Consumers’ response to the price reductions and promotions by competitive products may be changed by the introduction of loyalty programme of the focal product ([Leenheer, van Heerde et al. 2007](#_ENREF_42)). When consumers become loyalty programme members of a specific brand, they receive saving rewards and direct discounts, and may find the promotions of alternative brands less attractive. This also applies to the termination of existing loyalty program ([Melnyk and Bijmolt 2007](#_ENREF_50)).

Some studies tried to capture the changing process of the effectiveness of the marekting actitives. Foekens, S.H. Leeflang et al. ([1999](#_ENREF_28)) 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 modelled as a function of historical promotional information of the focal brand and other competitive brands. The model tries 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_38)) also proposed extensions of the SCAN\*PRO model to investigate the dynamic impact of promotions on the baseline sales. Their results indicate that promotions increase the concurrent product sales but reduce the baseline sales. These models, however, are all descriptive models and are not used in forecasting retailer product sales.

1. **Structural break in promotional models and forecast bias**

In this section, we briefly introduce the issue of structural break and its impact on the forecasting performance of conventional econometric models we use to forecast retailer product sales. When the effectiveness of the price and promotions on product sales change, as described in previous section, 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_3)). The parameter estimates of the models then become the weighted average of the true parameters before and after the structural break. The forecasts generated by the model will be subject to bias and less accurate[[1]](#footnote-1). The impact of structural break on the model’s forecasting performance has been addressed by many studies in the macroeconomics literature (e.g. [Cooper and Nelson 1975](#_ENREF_16), [Muellbauer 1994](#_ENREF_54), [Hendry 1995](#_ENREF_31), [Clements and Hendry 1999](#_ENREF_15), [Pesaran and Timmermann 2007](#_ENREF_60), [Castle, Doornik et al. 2008](#_ENREF_11)).

[Pesaran and Timmermann (2005)](#_ENREF_62) show some analytical analysis for the impact of a structural break on the forecasting performance of a simple regression model. For example, suppose that we have the sales and price data from week 1 to week *T,* i.e., and a structural break occurs at the date of (where ). We assume that the parameters for the price variable changes from to after . In practice, 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. which are unobservable to us. We assume that the real demand is of the following process:

where, is an indicator which equals to 1 before week and 0 otherwise. and are respectively the dependent variable and the explanatory variable at week *t*. is assumed to be strictly exogenous[[2]](#footnote-2). and are the parameters before and after the structural break at week . is the error term, and we assume . We also assume that the variance of the error term shifts from to after week .

We may estimate a model which is congruent with the demand (e.g., ). We denote that the model is estimated with an estimation window starting before the structural break, e.g., at week *m* . The OLS estimate for the model is therefore:

where and are the matrices for the explanatory variable and the dependent variable respectively with the observations from week *m* to week T. Since the true parameter for the price variable changes from to within the estimation period, cannot be an unbiased estimate of but a weighted average of the true parameters before and after the structural break. This example assumes that there is no structural break after week T, and the true demand after week T will remain as . Therefore, the h-step ahead forecast error at week *T*+h can be represented as:

where , is the matrix for the error term at week .

Accordingly, the forecast bias at week , which is , and is unequal to zero as .

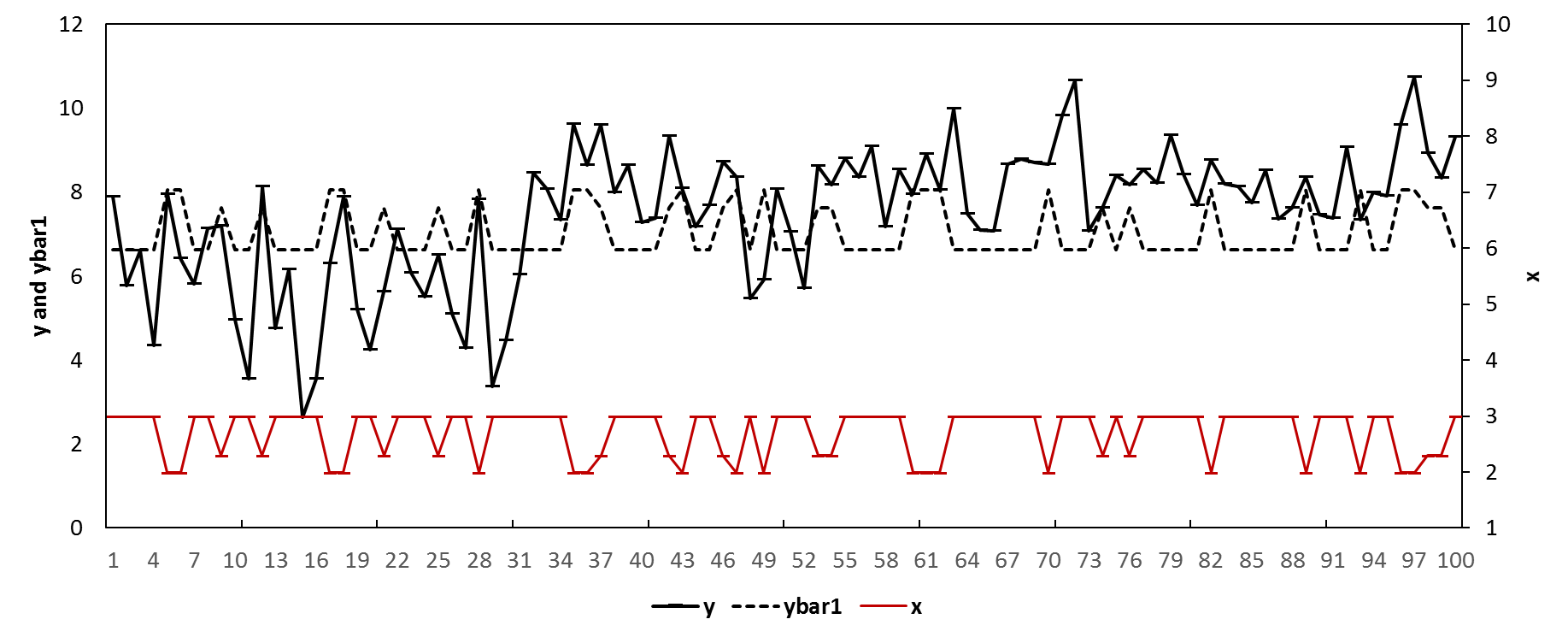
This can be further illustrated using a simulation. For example, we assume that the product price is usually 2.99 with occasional temporary price reductions to 2.29 or 1.99[[3]](#footnote-3). e.g., or 2.29 or 1.99. We also assume that the product sales are exclusively determined by the product price but with a structural break at week 31. The real demand is assumed as follows:

, , when

, , when

In this example, the sales generally increase but also become less responsive to temporary price reductions after the break. This could be typical for those products which enter from a growth stage to a mature stage regarding their product life cycle[[4]](#footnote-4). The sales and price data are represented in Figure 1 by the solid black line and the solid red line respectively.

Figure 1. Simulated sales with a structural break: model with full data



Suppose we have the data from week 1 to week 70 and we want to forecast the product sales from week 71 to week 100. In Figure 1, the blue area represents the estimation period before the structural break (e.g., from week 1 to week 30), the yellow area represents the estimation period after the structural break (e.g., from week 30 to week 70), and the red area represents the forecast period (e.g., from week 71 to week 100). We may estimate the model with the function form using the data from week 1 to week 70 and overlook the structural break at week 31. Under such circumstance, we will have estimates as the weighted average of the true parameters before and after week 31. We will over-predict the product sales for the period from week 1 to week 30 and under-predict the product sales for the period from week 31 to week 70, and we will produce downwards-biased forecasts for the period from week 71 to week 100. The predictions/forecasts are represented by the black dashed line in Figure 1. Table 1 shows the forecasting performance of this model regarding some error measures.

Ideally, if we observe the structural break at week 31, we may estimate the model exclusively using the data from week 31 to week 70 and generate unbiased forecasts. We represent those unbiased forecasts using the black dashed line in Figure 2. However, in a retailing context we do not know neither the existence nor the location (e.g., week 31 in this example) of the structural break as the influencing factors are unobservable. Also if the structural break occurs close to the forecast origin (e.g., at week 68) and there will not be enough post-break observations to estimate the model.

Figure 2. Simulated sales with a structural break: model with post-break data

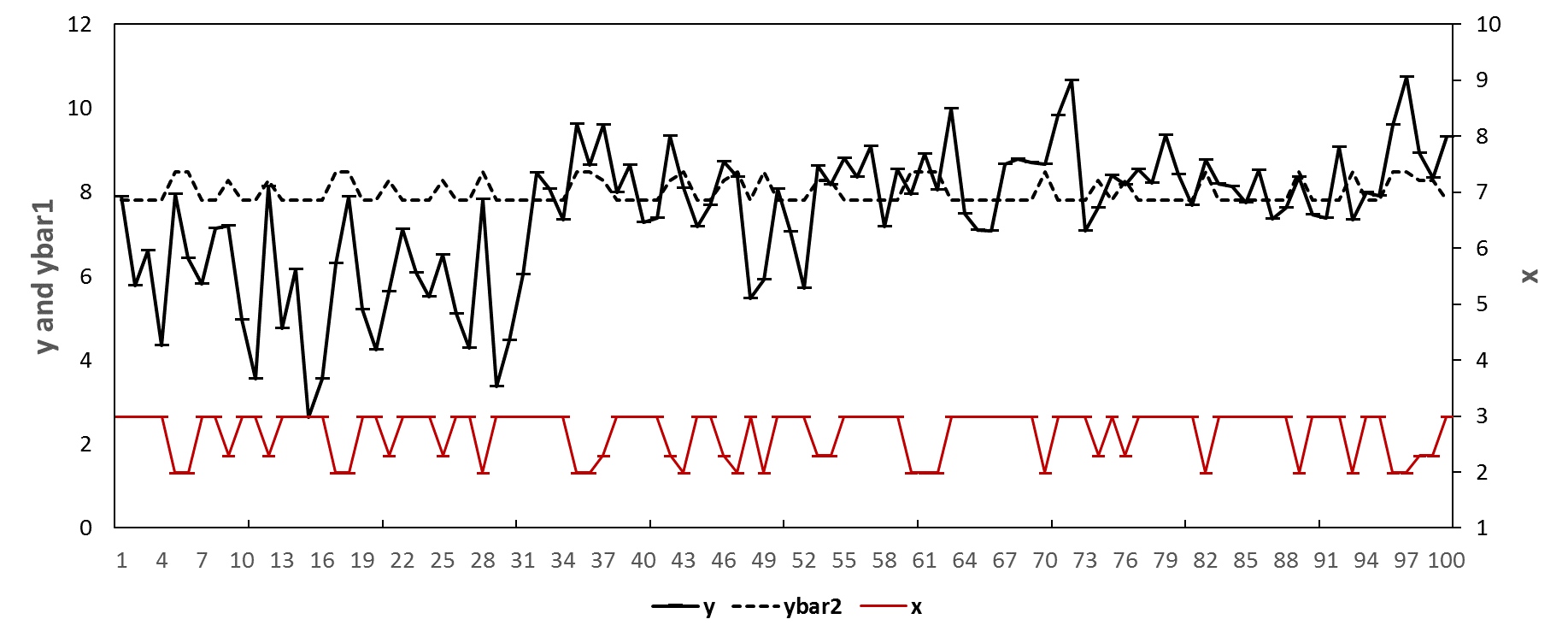


Figure 3.

Table 1. The forecasting performance of different models in the simulation

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MAE | MAPE | SMAPE |
| Figure 1: Model with full estimation window | 1.474 | 17.46% | 18.79% |
| Figure 2: Model with Post-break estimation window | 0.732 | 8.51% | 8.60% |
| Figure 3: Model with intercept correction | 0.824 | 9.77% | 9.54% |
| Figure 4: Model with estimation window combining | 1.034 | 12.17% | 12.58% |

1. **The methods**

4.1 Intercept correction

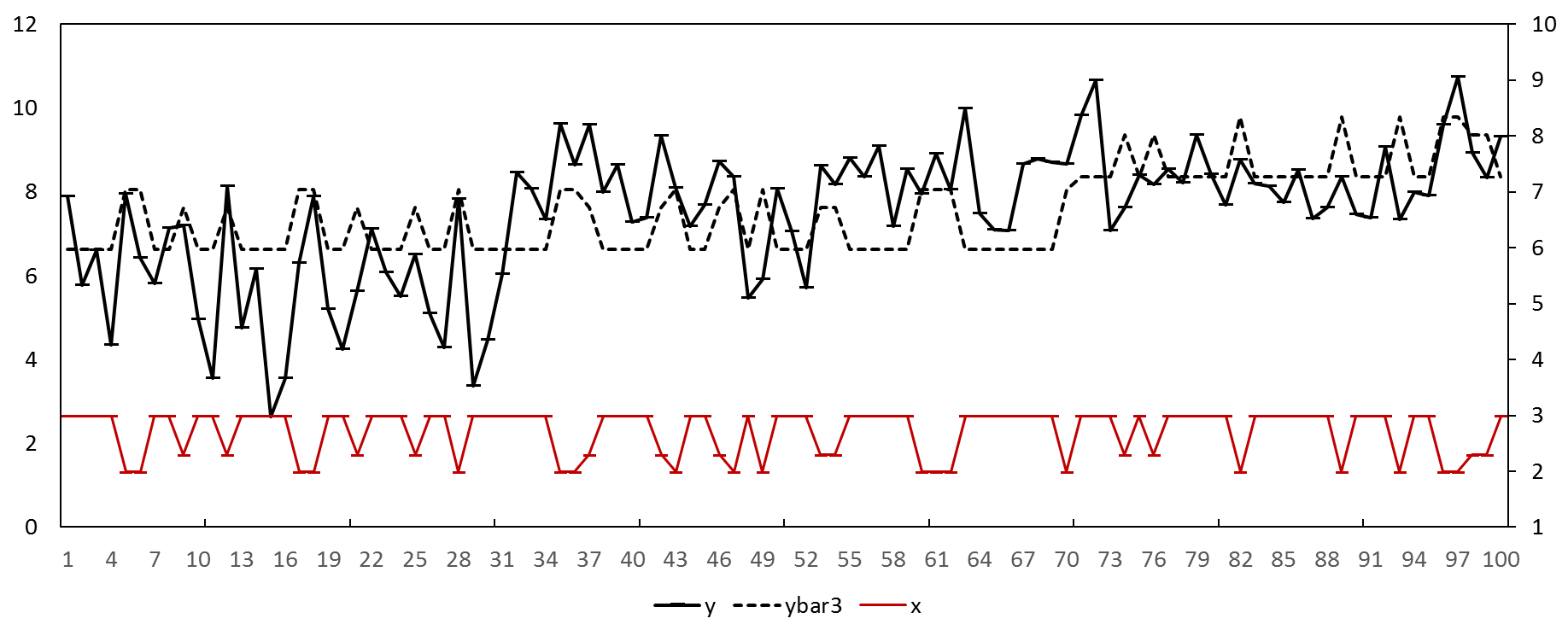
We can apply the intercept correction (IC) method to mitigate the bias contained in the forecasts when the model is subject to structural break. The IC method has been commonly applied in the macro-economic studies to offset regime shifts ([Clements and Hendry 1994](#_ENREF_14), [Clements and Hendry 1999](#_ENREF_15), [Clark and McCracken 2007](#_ENREF_12)). The method first detects the existence of the structural break. If the model is subject to structural break, the method tries to estimate the forecast bias and offset the forecast bias by specifying non-zero values for the model’s errors in the forecasting period. The intercept correction technique may potentially improve the forecasting accuracy by mitigating the forecast bias at the cost of inflated forecasting error variance ([Clements and Hendry 1999](#_ENREF_15)).

The IC method can be demonstrated using the same simulation described in section 3 where we have the model as but with no prior knowledge related to the location of the structural break. We first conduct a sequential Chow test based on most of the observations in the estimation period[[5]](#footnote-5). Figure 3 shows the *p*-values of the sequential Chow test. Figure 3 indicates the *p*-values of the Chow test assuming there is one single structural break occurring at a specific week. The results reject the null hypothesis of no structural break for the weeks which are from week 16 to week 63[[6]](#footnote-6). Although the sequential Chow test do not suggest the location of the structural break, it indicates whether structural break exists during the estimation period. In the literature, more advanced statistic tests have been proposed to detect the locations of the structural breaks but they all need to assume additional priori knowledge such as the number of potential multiple structural breaks ([Andrews 1993](#_ENREF_4), [Andrews and Ploberger 1994](#_ENREF_5), [Bai and Perron 2003](#_ENREF_9)).

Figure 3 P-values of the sequential Chow test

Based on the results in Figure 3, we consider that the model is subject to structural break and generate biased forecasts. We could estimate the magnitude of the forecast bias with different schemes. For example, we may estimate the bias as the predictive error at the forecast origin (i.e., , where *T* =70). Alternatively, we may robustly estimate the bias as the average value of an ad hoc number of predictive errors before the forecast origin. (e.g. , where *i* can be arbitrarily chosen). In this example, we calculate the predictive error for the last four observations in the estimation period. i.e., .

Figure 4. Simulated sales with a structural break: model with intercept correction



We then mitigate the forecast bias by adding bias estimate back to the forecasts. e.g., . The ‘intercept corrected’ forecasts are represented by the black dashed line in Figure 4. These forecasts are more accurate compared to those by the original model (e.g., MAE= 0.824, MAPE= 9.77%, and SMAPE= 9.54%). The forecast/predict value of the product sales are shown in Figure 4. The forecasting performance regarding various error measures are shown in Table 1.

One of the limitations for the intercept correction method is that it heavily relies on the detection and estimation of the forecasts bias. The estimation of the forecast bias can be very challenging if the data variation for the estimation period close to the forecast origin is very large. Also, the IC method adds the estimated bias back to the forecasts, which inevitably inflates the error variance of the forecasts. Evidence shows that the inflation of the forecasting error variance can be different for different correction schemes described above based on how the true parameters change and whether there are multiple structural break (Clements and Hendry ([1999](#_ENREF_15)). Finally, the IC method assumes that there is no structural break during the time period close to the forecast origin (e.g., in the simulation example, we presume that the structural break does not occur within the last four observations in the estimation window). Therefore, whether we can generate more accurate forecasts by implementing the IC method to conventional models for retailer product sales at the SKU level becomes an empirical question.

4.2 Estimation window combining

The intercept correction method may potentially improve the forecasting performance of conventional models, but relies on how accurate the forecast bias can be estimated. An alternative method which circumvents the estimation of forecast bias is to combine the forecasts generated by the same model but with different estimation windows. As described in section 3, if we know the location of the structural break, we could estimate the model exclusively with the data after the structural break (e.g., from week 31 to week 70). Under this condition, the model will not be subject to structural break and generates unbiased forecasts, as shown in Figure 2. However, in reality, the location of the structural break is not observable. It is also difficult to estimate the location and the size of the structural break ([Pesaran and Timmermann 2007](#_ENREF_60)). We may try to exclude the data before the structural break by estimating the model with only the most recent observations close to the forecast origin and keeping the size of the estimation window small as long as there are enough observations to estimate the model. It is reasonable to do so because we may assume that with fewer (and more recent) observations in the estimation, the model will be less likely to be subject to structural break and be less likely to produce biased forecasts. For the model in section 3, we may keep *m* as large as possible (so that we have a small estimation window). when *m* becomes larger than , the model will not be subject to structural break and will generate unbiased forecasts. For the same simulation example in section 3, we may arbitrarily choose to estimate the model using the data from week 50 to week 70 even when we do not observe the date of the structural break. The forecasts generated under such circumstance will be unbiased (or least less biased).

However, the generated unbiased or less biased forecasts may not necessarily be more accurate because of the associated cost: when we estimate the model using a smaller estimation window compared to the original full estimation window (e.g., the data from week 50 to week 70), we drop off part of the information, which, as a result, leads to an increase of the forecasting error variance. This can be demonstrated with the same example used in section 3. For example, the forecast error is represented as follows:

We may consider the condition when the error measure is Mean Square Forecasting Error (MSFE) as in [Pesaran and Timmermann (2007)](#_ENREF_60). The h-step ahead forecast conditional on is represented as:

where

is interpreted as the squared forecast bias, and is interpreted as the efficiency term ( is the forecasting error variance). We may re-estimate the model using one additional observations before the structural break to investigate the change of the MSFE. Thus the change of the MSFE’s for week when we include one more observation in the model is shown as:

where is the MSFE for the model which is estimated with the data from week m-1 to week T (the estimation window from week *T*-m to week *T*). It can be proved that the bias term () (i.e., the change of the squared forecast bias) is always larger than or equal to zero (i.e., with one more observation before the structural break, the forecast gets more biased), but the sign for the efficiency term depends on the percentage of the change in the error variance before and after the structural break compared to the error variance after the structural break (i.e., ). If (e.g., there are more post-break variations in the product sales which cannot be explained by the price variable), will be larger than or equal to , and the MSFE may rise as both terms are non-negative. However, if (e.g., there are less variations in the product sales which cannot be explained by the price variable), will be smaller than or equal to . Under this circumstance, the MSFE will either rise or fall depending on how the non-negative squared bias term compares to the non-positive efficiency term. Therefore, when we add observations before the structural break, we may have either better or worse forecasting performance depending on the retailer sales data due to the trade-off between the increased forecast bias and the potentially reduced forecasting error variance.

Under such circumstance, we may combine the forecasts generated by the models with different estimation windows. [Pesaran and Timmermann (2007)](#_ENREF_60) introduced the combination schemes based on the cross-validation MSFE and equal weights. In this study, we implement the combination scheme with equal weights because it has been proved with good performance and easy to implement.([Clements and Hendry 1998](#_ENREF_13), [Fildes and Stekler 2002](#_ENREF_26), [Dekker, van Donselaar et al. 2004](#_ENREF_22), [Pesaran, Schuermann et al. 2009](#_ENREF_61)). Specifically, we can estimate the model using the most recent observations (i.e. the data from week to ) to generate the first set of h-step-ahead forecast as:

where 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. We can the repeat this process by adding more observations to the estimation window and generate forecasts. For example, we may have the the set of h-step-ahead forecast as:

Eventually, we calculate the average of the () sets of h-step-ahead forecasts based on equal weights:

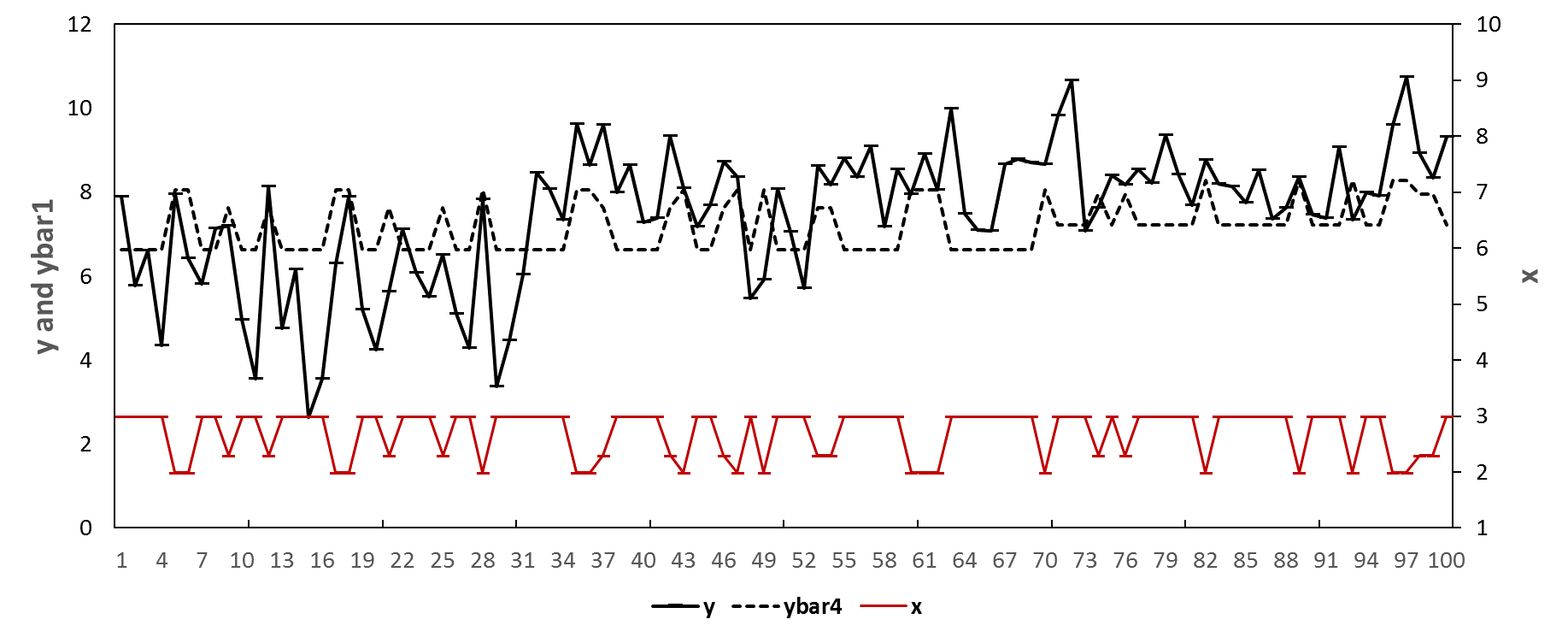
This can be illustrated using the same simulation example in section 3. Suppose that we know that there is a structural break within the estimation period but we do not know the location of the break (e.g., we do not know it is from week 31), then we may choose to use the full estimation window (e.g., all the available data, from week 1 to week 70) to estimate the model. Under this situation, the forecasts will be subject to the full bias, as illustrated in section 3. Alternatively, we may estimate the model with only the latest data (e.g., the data period from week 41 to week 70)[[7]](#footnote-7). Under this situation, it is less likely for the forecasts to be subject to bias. However, this leads to the coast of inflated forecasting error variance because of less information used in the estimation (e.g., only 30 observations are used in the estimation, compared to 70 observations used in the former scenario).

Under such condition, the forecasts generated by some models (e.g., estimated with the full data, as the one in Figure 1) will be biased, and the forecasts generated by other models (e.g., estimated with the data after the structural break, as the one in Figure 2) will be less biased (and unbiased if all the data used in the estimation are all post-break data). If we combine the forecasts generated by these models, the forecasts will be less biased compared to the forecasts generated in Figure 1. However, there is cost of the inflated forecasting error. When we generate the forecasts in Figure 2 using the post-break data, we are discarding the data before the break which may still contains useful information for the relationship between the independent variables and the dependent variable. With a smaller estimation window, we tend to have larger forecasting error variance. Therefore, with the estimation window combing method, we may expect to have more accurate forecast results from a better trade-off between the inflated forecast error variance and the reduced forecast bias which both contribute to the forecasting performance.

In the simulation example, we estimate the model with different lengths of estimation windows. We estimate the model using the data from week 1 to week 70, and then generate the forecasts for the period after week 70. We denote this set of forecasts as . This set of forecasts are subject to the full bias. We then estimate the same model but using the data from week 2 to week 70, and we generate forecasts for the period after week 70 and denote them as , and so forth, and these forecasts will be less biased compared to . In this simulation, we generate 40 sets of forecasts using estimation windows from [1:70] to [40:70]. We may choose an arbitrary number of sets as long as there are enough observations to estimate the model even with the smallest estimation window. At last, we combine these 40 sets of forecasts using equal weight average. i.e.,. The forecasts, which are illustrated by the black dashed line in Figure 5, are more accurate compared to the forecasts by the original model with the error measures shown in Table 1. (e.g., 1.034 for MAE, 12.17% for MAPE, and 12.58% for SMAPE).

The EWC method relies on the trade-off between the reduced forecast bias and the inflated forecast error variance. In this study, we evaluate the empirical question that whether we can generate more accurate forecasts by implementing the EWC method to conventional models for retailer product sales at the SKU level.

Figure 5. Simulated sales with a structural break: model with estimation window combing



1. **The data**

In this study, we propose more effective models to retailer product sales at the SKU level. We evaluate our models using the dataset made available by the IRI company. A descriptive article can be found in [Bronnenberg, Kruger et al. (2008)](#_ENREF_10)[[8]](#footnote-8). The dataset contains weekly data at the SKU level with variables including unit sales, price, features and displays etc. for more than seven years[[9]](#footnote-9). We randomly select 128 SKUs with positive movements for at least 90% of time in 15 product categories in a large store. Table 2 shows some basic statistics for the selected SKU’s in each of the product categories. The table indicates that some product categories (e.g., Carbonated beverages and Hotdog) have much higher promotional intensity compared to others (e.g., Margarine/Butter and Mayonnaise). Figure 6 depicts the sales data for a typical SKU in the Beer category. The product is not intensively promoted only with occasional price reductions and feature/display events where the product sales exhibits spikes accordingly.

Table 2. The statistics for the SKUs in each of the product categories

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Number of SKUs | Sales mean | Sales std | Price mean | Price std | Feature percentage | Display percentage |
| Beer | 9 | 30 | 11 | 7.30 | 0.39 | 0.91% | 2.72% |
| Carbonated beverages | 9 | 281 | 403 | 2.28 | 0.38 | 26.71% | 55.18% |
| Coffee | 9 | 41 | 37 | 4.66 | 0.62 | 6.57% | 15.28% |
| Frozen pizza | 8 | 35 | 39 | 3.20 | 0.32 | 7.33% | 7.87% |
| Household cleaner | 9 | 30 | 10 | 2.47 | 0.11 | 2.51% | 0.11% |
| Hotdog | 9 | 105 | 163 | 4.33 | 0.81 | 18.80% | 18.38% |
| Laundry detergent | 9 | 51 | 105 | 6.32 | 0.71 | 10.15% | 3.95% |
| Margarine/Butter | 9 | 107 | 101 | 2.35 | 0.25 | 8.12% | 0.05% |
| Mayonnaise | 9 | 43 | 13 | 2.54 | 0.17 | 0.32% | 0.00% |
| Mustard & ketchup | 7 | 36 | 29 | 2.85 | 0.27 | 0.27% | 0.69% |
| Peanut butter | 8 | 31 | 16 | 3.50 | 0.36 | 0.60% | 4.09% |
| Salty snacks | 9 | 56 | 71 | 2.77 | 0.27 | 6.57% | 4.86% |
| Soup | 9 | 184 | 239 | 1.22 | 0.14 | 9.72% | 2.72% |
| Sugar substitutes | 6 | 12 | 6 | 2.59 | 0.21 | 1.12% | 0.00% |
| Toothpaste | 9 | 31 | 58 | 2.59 | 0.27 | 11.32% | 11.81% |

Figure 6. Unit sales, price (in USD), and promotional events (feature and display) for an SKU in the Beer category.



1. **Models**

We include the base-lift method as one of the benchmark models as it has been widely used by retailers to forecast product sales at the SKU level. The method generates baseline forecasts using simple univariate models (e.g., the simple exponential smoothing model) and then makes adjustments for any incoming promotional event. For example,

Where is the baseline forecast for week . is the actual sales at the previous week when the focal product is not being promoted. is the parameter for the simple exponential smoothing model. The adjustment is calculated as the increased sales by the most recent promotional event for the focal product.

In this study, we propose to improve the performance of the conventional forecasting models by accounting for the forecast bias caused by structural break. We first evaluate whether we could improve the performance of the ADL-own models proposed by Huang et al. (2014) which is a sophisticated Autoregressive Distributed Lag (ADL) model with price and promotional information of the focal product (i.e., referred as the ADL-own model) simplified by the general-to-specific strategy. The advantage of this model is that it captures the dynamic effects of the price and promotional variables with a parsimonious specification. In practice, manufacturers may not always be able to access to retailers’ sales information ([Ali and Boylan 2011](#_ENREF_2)). Under this circumstance, manufacturers need to make the best use of the data they have and produce forecasts as accurate as possible. The ADL-own model shall be initially constructed with sophisticated dynamic terms but only based on limited information. For example, the general model form is:

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 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 *[[10]](#footnote-10)*

are the parameters  
 is the error term and we assume

is the order of the lags[[11]](#footnote-11).

The model is then tested downwards with one variable at a time and eventually becomes the valid and most parsimonious form of the general model ([see Huang, Fildes et al. 2014](#_ENREF_34)). The ADL-own model assumes that the effect of the price and promotional activity of the focal product to be constant over time and may be subject to structural break. In this study, we implement the intercept correction method and the estimation window combining method on the ADL-own model. We refer the new models as the ADL-own-IC model and the ADL-own-EWC model respectively. These two new models may potentially generate more accurate forecasts compared to the ADL-own model, which is particularly helpful for the manufacturers who do not get access to competitive information from the retailer and rely on exclusively their own promotional data.

We first evaluate whether we could improve the performance of the ADL model and the ADL-DI models. These two models outperformed the ADL-own model by incorporating the competitive price and competitive promotional information within the same product category, where the competitive price and promotional information were integrated through a refining process based on variable selection methods (e.g., the LASSO algorithm) and principle component analysis (i.e., PCA). The specifications for the general ADL model and the general ADL-DI model is shown as follows:

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

are the parameters  
 is the error term and we assume

is the order of the lags.

where

is the diffusion index of competitive price at week .

is the diffusion index of competitive promotion at week .

*P* and *Q* are the number of initially retained diffusion indexes, and

In this study, we implement the intercept correction method and the estimation window combining method on the ADL model and the ADL-DI model respectively. We refer the new models with intercept correction as the ADL-IC model and the ADL-DI-IC model, and we refer the new models with estimation window combining as the ADL-EWC model and the ADL-DI-EWC model. These new models may potentially generate more accurate forecasts for retailers who get access to all competitive promotional information within the same product category.

In this study, the intercept correction method and the estimation window combining method are both implemented discriminately based on the results of the sequential Chow test for the existence of structural break[[12]](#footnote-12), as described in section 3. For the intercept correction method, we estimate the forecast bias based on equally weighted average of four predictive errors before the forecast origin, and we make direct adjustments to the *h*-step-ahead forecasts using the full amount of the estimate for the forecast bias. There can be different correction schemes for the forecast bias when the model contained lagged product sales as explanatory variables. Different correction schemes lead to slightly different characteristics regarding bias reduction and forecast error variance inflation ([Clements and Hendry 1999](#_ENREF_15)). In this study, we make adjustments directly to the *h*-step-ahead forecast using the full amount of the forecast bias. An alternative scheme 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 already adjusted, and so forth. For the estimation window combining method, we adopt a combining scheme of equal weights as equal weighting scheme has been proved to be effective and easy to implement. The candidate models are summarized in Table 3:

Table 3. The candidate models with description

|  |  |
| --- | --- |
| Models | Description |
| Base-lift | Industrial practice, Simple-exponential smoothing with adjustments based on the effect of the most recent promotional event |
| ADL-own | Generate-to-specific ADL model, with promotional variables of the focal product only |
| ADL | Generate-to-specific ADL model, based on the variables retained by LASSO |
| ADL-own-EWC | ADL-own model, with EWC |
| ADL-own-IC | ADL-own model, with IC |
| ADL-IC | ADL model with IC |
| ADL-EWC | ADL model with EWC |
| ADL-DI | Generate-to-specific ADL model, based on the diffusion index constructed by principle component analysis |
| ADL-DI-IC | ADL-DI model, with IC |
| ADL-DI-EWC | ADL-DI model, with EWC |

1. **The experimental design**

In this study, we evaluate the forecasting performance of the models with both fixed origin and rolling origins ([Tashman 2000](#_ENREF_65)). For the setting with the fixed origin, we specify the models with the data from week 1 to week 150 and generate the forecasts for one to weeks ahead, where is 1, 4, and 12 which are typically the ordering and planning period for retailers. For the setting with rolling origins, we start with an estimation window from week 1 to week 120 and generate one to weeks ahead forecasts. We then re-estimate the model with updated data for the following week and without the data for the earliest week (e.g., from week 2 to week 121), and so forth. In this study, we conduct 20 sets of rolling experiments and generate 20 sets of one to weeks ahead forecast in total. We use the actual values of the exogenous variables (e.g., price, promotion, or calendar events etc.) and the forecasts of the lagged dependent variables when the forecast horizon is beyond one week. For the setting with rolling origins, we specify the ADL models with the data from week 1 to week 150 assuming a foreknowledge of the data ([Fildes, Wei et al. 2011](#_ENREF_27)). An alternative way is to re-specify the ADL models for each rolling estimation window ([Ma, Fildes et al. 2016](#_ENREF_47)). With rolling forecast origins, the results are more robust to randomness and systematic business cycle effects ([Fildes 1992](#_ENREF_23), [Stock and Watson 2002](#_ENREF_64), [Stock and Watson 2002](#_ENREF_63)).

We evaluate the models’ forecasting performance using four error measures: the Mean Absolute Percentage Error (MAPE), the symmetric Mean Absolute Percentage Error (sMAPE), the Mean Absolute Scaled Error (MASE) ([Hyndman and Koehler 2006](#_ENREF_35)), and the Relative Average Mean Absolute Error (RelAvgMAE) ([Davydenko and Fildes 2013](#_ENREF_20)). These error measures approximate the loss function of the retailer from different aspects. The error measures for SKUs and rolling events based on forecast horizon of 1 to (i.e. , , and =1, 4 and 12) are as follows:

where is the actual value in the forecast period for data series based on the rolling event. is the forecast value for data series based on the rolling event[[13]](#footnote-13). is the total number of observations in the full estimation window.

1. **Results and discussion**
   1. fixed origin

Table 4a shows the forecasting performance of the various models with a fixed forecast origin. The Base-lift model are outperformed by all the other candidate models in almost all the scenarios. The ADL-own model is always outperformed by the ADL model, and has mixed forecasting performance compared to the ADL-DI model. The ADL-own model gets outperformed by the ADL-own-IC model and the ADL-own-EWC for all the scenarios. The ADL model is outperformed by the ADL-IC model and the ADL-EWC model for almost all the forecast horizons and error measures. The ADL-DI model is always outperformed by the ADL-DI-IC model. The ADL-DI-EWC model have mixed results compared to the ADL-DI model. Overall, the ADL-IC model and the ADL-EWC model generally have the best forecasting performance.

Table 4a. The forecasting performance of various candidate models with a fixed forecasting origin

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fixed origin, Forecast horizon= 1 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-lift | 39.60% | 7 | 37.58% | 10 | 0.396 | 7 | 1.000 | 6 |
| ADL-own | 42.24% | 10 | 36.65% | 9 | 0.422 | 10 | 1.077 | 9 |
| ADL | 38.74% | 6 | 33.48% | 3 | 0.387 | 6 | 0.953 | 2 |
| ADL-own-EWC | 41.17% | 9 | 36.00% | 8 | 0.412 | 9 | 1.096 | 10 |
| ADL-own-IC | 39.69% | 8 | 34.62% | 7 | 0.397 | 8 | 1.069 | 8 |
| ADL-IC | 37.48% | 5 | 32.63% | 2 | 0.375 | 5 | 0.999 | 5 |
| ADL-EWC | 37.41% | 4 | 32.58% | 1 | 0.374 | 4 | 0.906 | 1 |
| ADL-DI | 36.97% | 3 | 34.53% | 6 | 0.370 | 3 | 1.025 | 7 |
| ADL-DI-IC | 35.98% | 2 | 33.53% | 4 | 0.360 | 2 | 0.961 | 3 |
| ADL-DI-EWC | 35.82% | 1 | 33.63% | 5 | 0.358 | 1 | 0.986 | 4 |
| Fixed origin, Forecast horizon= 4 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-lift | 41.12% | 10 | 37.98% | 10 | 0.776 | 10 | 1.000 | 10 |
| ADL-own | 36.97% | 6 | 34.42% | 6 | 0.692 | 6 | 0.906 | 6 |
| ADL | 36.17% | 4 | 33.88% | 4 | 0.674 | 4 | 0.895 | 4 |
| ADL-own-EWC | 36.35% | 5 | 34.39% | 5 | 0.690 | 5 | 0.902 | 5 |
| ADL-own-IC | 35.66% | 3 | 33.28% | 1 | 0.673 | 3 | 0.877 | 2 |
| ADL-IC | 35.64% | 2 | 33.36% | 2 | 0.667 | 1 | 0.883 | 3 |
| ADL-EWC | 35.49% | 1 | 33.71% | 3 | 0.669 | 2 | 0.876 | 1 |
| ADL-DI | 38.21% | 8 | 35.48% | 8 | 0.755 | 8 | 0.925 | 8 |
| ADL-DI-IC | 37.55% | 7 | 35.00% | 7 | 0.744 | 7 | 0.922 | 7 |
| ADL-DI-EWC | 38.31% | 9 | 35.60% | 9 | 0.766 | 9 | 0.931 | 9 |
| Fixed origin, Forecast horizon= 12 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-lift | 44.72% | 10 | 40.94% | 10 | 0.846 | 10 | 1.000 | 10 |
| ADL-own | 37.87% | 6 | 35.48% | 6 | 0.762 | 5 | 0.903 | 6 |
| ADL | 36.40% | 3 | 34.43% | 3 | 0.725 | 2 | 0.864 | 3 |
| ADL-own-EWC | 37.10% | 5 | 35.45% | 5 | 0.769 | 6 | 0.897 | 5 |
| ADL-own-IC | 36.47% | 4 | 34.35% | 2 | 0.744 | 4 | 0.870 | 4 |
| ADL-IC | 35.99% | 2 | 34.08% | 1 | 0.721 | 1 | 0.860 | 1 |
| ADL-EWC | 35.86% | 1 | 34.49% | 4 | 0.731 | 3 | 0.860 | 2 |
| ADL-DI | 38.72% | 9 | 36.47% | 8 | 0.782 | 8 | 0.915 | 9 |
| ADL-DI-IC | 38.19% | 8 | 35.91% | 7 | 0.772 | 7 | 0.907 | 7 |
| ADL-DI-EWC | 38.08% | 7 | 36.49% | 9 | 0.791 | 9 | 0.912 | 8 |

8.2 rolling origins

Table 1b shows the forecasting performance of the various regarding rolling forecast origins. The Base-lift model are now outperformed by all the other candidate models in all the scenarios. The ADL-own model is outperformed by the ADL model and the ADL-DI model for all the forecast horizons for all the error measures. The results are consistent with Huang et al. (2014) and Ma et al. (2016). The ADL-own model is again outperformed by the ADL-own-IC model and the ADL-own-EWC model for almost all the scenarios. The results confirm that we general more accurate forecasts using the IC method and the EWC method for manufacturers when competitive information are not available. The ADL model is again outperformed by the ADL-IC model for all the forecast horizons and for all the error measures. The ADL-EWC model has mixed forecasting performance compared to the ADL model. The ADL-DI model gets outperformed by the ADL-DI-IC model all the scenarios. The ADL-DI-EWC model has mixed forecasting performance compared to the ADL-DI model. Overall, the ADL-IC model and the ADL-DI-IC model generally have the best forecasting performance.

Table 4b. The forecasting performance of various candidate models with rolling forecast origins

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Rolling origin, Forecast horizon= 1 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-lift | 42.80% | 10 | 38.55% | 10 | 0.799 | 10 | 1.000 | 10 |
| ADL-own | 34.34% | 9 | 30.57% | 8 | 0.669 | 8 | 0.898 | 9 |
| ADL | 30.88% | 2 | 29.33% | 3 | 0.649 | 4 | 0.855 | 4 |
| ADL-own-EWC | 34.23% | 8 | 30.66% | 9 | 0.674 | 9 | 0.893 | 8 |
| ADL-own-IC | 33.45% | 7 | 29.82% | 7 | 0.654 | 6 | 0.866 | 6 |
| ADL-IC | 30.07% | 1 | 28.78% | 1 | 0.642 | 2 | 0.833 | 1 |
| ADL-EWC | 30.95% | 3 | 29.55% | 4 | 0.661 | 7 | 0.868 | 7 |
| ADL-DI | 31.94% | 5 | 29.62% | 5 | 0.644 | 3 | 0.858 | 5 |
| ADL-DI-IC | 31.58% | 4 | 29.30% | 2 | 0.636 | 1 | 0.848 | 2 |
| ADL-DI-EWC | 32.47% | 6 | 29.73% | 6 | 0.654 | 5 | 0.852 | 3 |
| Rolling origin, Forecast horizon= 4 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-lift | 43.61% | 10 | 39.36% | 10 | 0.800 | 10 | 1.000 | 10 |
| ADL-own | 34.37% | 9 | 31.66% | 9 | 0.676 | 8 | 0.842 | 9 |
| ADL | 31.86% | 6 | 30.19% | 5 | 0.657 | 6 | 0.802 | 5 |
| ADL-own-EWC | 33.94% | 8 | 31.63% | 8 | 0.681 | 9 | 0.838 | 8 |
| ADL-own-IC | 32.78% | 7 | 30.34% | 7 | 0.652 | 5 | 0.804 | 6 |
| ADL-IC | 30.42% | 1 | 29.15% | 1 | 0.643 | 3 | 0.772 | 3 |
| ADL-EWC | 31.61% | 5 | 30.24% | 6 | 0.663 | 7 | 0.804 | 7 |
| ADL-DI | 30.85% | 3 | 29.91% | 3 | 0.637 | 2 | 0.773 | 4 |
| ADL-DI-IC | 30.44% | 2 | 29.45% | 2 | 0.628 | 1 | 0.760 | 1 |
| ADL-DI-EWC | 30.95% | 4 | 29.94% | 4 | 0.644 | 4 | 0.772 | 2 |
| Rolling origin, Forecast horizon= 12 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-lift | 48.82% | 10 | 41.38% | 10 | 0.851 | 10 | 1.000 | 10 |
| ADL-own | 37.17% | 9 | 33.21% | 9 | 0.725 | 9 | 0.815 | 9 |
| ADL | 33.85% | 6 | 31.40% | 6 | 0.696 | 6 | 0.771 | 5 |
| ADL-own-EWC | 36.48% | 8 | 33.03% | 8 | 0.721 | 8 | 0.813 | 8 |
| ADL-own-IC | 35.40% | 7 | 31.54% | 7 | 0.693 | 5 | 0.776 | 7 |
| ADL-IC | 32.23% | 2 | 30.15% | 1 | 0.680 | 4 | 0.744 | 4 |
| ADL-EWC | 33.55% | 5 | 31.39% | 5 | 0.700 | 7 | 0.773 | 6 |
| ADL-DI | 32.27% | 3 | 31.02% | 3 | 0.664 | 2 | 0.737 | 3 |
| ADL-DI-IC | 31.96% | 1 | 30.58% | 2 | 0.655 | 1 | 0.724 | 1 |
| ADL-DI-EWC | 32.30% | 4 | 31.06% | 4 | 0.668 | 3 | 0.737 | 2 |

8.3 Forecasting performance regarding promotions on the focal product

We investigate the forecasting performance of the various models regarding whether the focal product is being promoted for rolling forecast origins[[14]](#footnote-14). Table 5a shows the results for the time period when the focal product is not being promoted. The base-lift method is outperformed by all the other models in almost all scenarios. The ADL-own model, the ADL model, and the ADL-DI model are respectively outperformed by the ADL-own-IC method, the ADL-IC method, and the ADL-DI-IC method for all the forecast horizons and for all error measures. The ADL-own-EWC model gets outperformed by the ADL-own model for one week ahead forecast horizon, but outperforms the ADL-own model for forecast horizons of one to four weeks and one to twelve weeks. The forecasting performance of the ADL-EWC model compared to the ADL model is also mixed. The ADL-DI-EWC model however gets outperformed by the ADL-DI model for all the forecast horizons and for all the error measures. The results are mostly consistent with the forecasting performance of the various models regarding all the forecast period. Overall, the ADL-IC model and the ADL-DI-IC model have the best forecasting performance for the time period when the focal product is not being promoted.

Table 5a. The forecasting performance of various candidate models for the non-promoted period

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Rolling origins, the non-promoted period, Forecast horizon= 1 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-lift | 41.81% | 10 | 31.66% | 10 | 0.547 | 10 |
| ADL-own | 38.28% | 8 | 30.67% | 8 | 0.546 | 8 |
| ADL | 31.61% | 3 | 29.67% | 4 | 0.532 | 6 |
| ADL-own-EWC | 38.64% | 9 | 30.71% | 9 | 0.546 | 9 |
| ADL-own-IC | 36.50% | 6 | 29.73% | 6 | 0.527 | 4 |
| ADL-IC | 30.74% | 1 | 29.06% | 1 | 0.523 | 2 |
| ADL-EWC | 31.55% | 2 | 29.78% | 7 | 0.537 | 7 |
| ADL-DI | 36.36% | 5 | 29.60% | 3 | 0.525 | 3 |
| ADL-DI-IC | 35.54% | 4 | 29.24% | 2 | 0.516 | 1 |
| ADL-DI-EWC | 37.85% | 7 | 29.72% | 5 | 0.529 | 5 |
| Rolling origins, the non-promoted period, Forecast horizon= 4 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-lift | 45.46% | 10 | 32.51% | 10 | 0.568 | 8 |
| ADL-own | 35.29% | 9 | 31.87% | 9 | 0.577 | 10 |
| ADL | 32.39% | 6 | 30.52% | 7 | 0.559 | 6 |
| ADL-own-EWC | 34.96% | 8 | 31.72% | 8 | 0.576 | 9 |
| ADL-own-IC | 33.08% | 7 | 30.32% | 5 | 0.550 | 5 |
| ADL-IC | 30.75% | 1 | 29.33% | 2 | 0.543 | 3 |
| ADL-EWC | 32.01% | 4 | 30.44% | 6 | 0.561 | 7 |
| ADL-DI | 31.74% | 3 | 29.86% | 3 | 0.543 | 2 |
| ADL-DI-IC | 31.10% | 2 | 29.32% | 1 | 0.532 | 1 |
| ADL-DI-EWC | 32.05% | 5 | 29.86% | 4 | 0.546 | 4 |
| Rolling origins, the non-promoted period, Forecast horizon= 12 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-lift | 57.45% | 10 | 35.18% | 10 | 0.621 | 10 |
| ADL-own | 36.90% | 9 | 33.31% | 9 | 0.613 | 9 |
| ADL | 34.13% | 6 | 31.46% | 7 | 0.591 | 6 |
| ADL-own-EWC | 36.32% | 8 | 33.07% | 8 | 0.607 | 8 |
| ADL-own-IC | 34.44% | 7 | 31.45% | 6 | 0.579 | 4 |
| ADL-IC | 32.24% | 1 | 29.99% | 1 | 0.572 | 2 |
| ADL-EWC | 33.70% | 5 | 31.37% | 3 | 0.594 | 7 |
| ADL-DI | 32.70% | 3 | 31.38% | 4 | 0.576 | 3 |
| ADL-DI-IC | 32.27% | 2 | 30.87% | 2 | 0.567 | 1 |
| ADL-DI-EWC | 32.79% | 4 | 31.44% | 5 | 0.580 | 5 |

Table 5a shows the results for the time period when the focal product is being promoted. The base-lift method is again outperformed by all the other models in almost all scenarios. The ADL-own-IC model, the ADL-IC model, and the ADL-DI-IC model all have mixed forecasting performance compared to their corresponding models without intercept corrections, and they generally have better forecasting performance for longer forecast horizons.

The ADL-own-EWC model gets outperformed by the ADL-own model for one week ahead forecast horizon, but outperforms the ADL-own model for forecast horizons of one to four weeks and one to twelve weeks. The forecasting performance of the ADL-EWC model compared to the ADL model is also mixed. The ADL-DI-EWC model however gets outperformed by the ADL-DI model for all the forecast horizons and for all the error measures. The results are mostly consistent with the forecasting performance of the various models regarding all the forecast period. Overall, the ADL-IC model and the ADL-DI-IC model have the best forecasting performance for the time period when the focal product is not being promoted. The ADL-own-EWC model has comparative forecasting performance with the ADL-own model. The ADL-EWC model gets outperformed by the ADL model for all the scenarios. The ADL-DI-EWC model also has mixed comparative performance with the ADL-DI model. The results indicate that the improvement by implementing the IC method and the EWC method is moderate for the promoted period. One possibility could be that the potential improved forecasting performance due to a reduction of forecast bias has been immerged in the data variation during the promoted period when the sales of the focal product are very high.

Table 5b. The forecasting performance of various candidate models for the promoted period

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Forecast horizon= 1 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-lift | 65.27% | 10 | 98.27% | 10 | 3.235 | 10 |
| ADL-own | 39.35% | 8 | 37.38% | 7 | 2.025 | 4 |
| ADL | 34.65% | 1 | 35.65% | 1 | 1.997 | 3 |
| ADL-own-EWC | 39.01% | 7 | 38.33% | 9 | 2.101 | 9 |
| ADL-own-IC | 42.83% | 9 | 38.29% | 8 | 2.056 | 7 |
| ADL-IC | 35.67% | 5 | 36.42% | 2 | 2.032 | 5 |
| ADL-EWC | 35.73% | 6 | 37.03% | 6 | 2.088 | 8 |
| ADL-DI | 34.80% | 2 | 37.02% | 5 | 1.983 | 2 |
| ADL-DI-IC | 35.17% | 4 | 36.66% | 3 | 1.964 | 1 |
| ADL-DI-EWC | 35.04% | 3 | 36.89% | 4 | 2.033 | 6 |
| Forecast horizon= 4 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-lift | 67.03% | 10 | 98.45% | 10 | 2.896 | 10 |
| ADL-own | 38.23% | 8 | 36.34% | 6 | 1.624 | 4 |
| ADL | 34.60% | 4 | 35.55% | 2 | 1.672 | 6 |
| ADL-own-EWC | 37.37% | 7 | 37.03% | 9 | 1.689 | 7 |
| ADL-own-IC | 40.39% | 9 | 36.66% | 8 | 1.648 | 5 |
| ADL-IC | 34.89% | 5 | 35.82% | 5 | 1.700 | 8 |
| ADL-EWC | 34.99% | 6 | 36.51% | 7 | 1.727 | 9 |
| ADL-DI | 33.56% | 2 | 35.65% | 4 | 1.564 | 2 |
| ADL-DI-IC | 34.12% | 3 | 35.61% | 3 | 1.562 | 1 |
| ADL-DI-EWC | 33.42% | 1 | 35.38% | 1 | 1.599 | 3 |
| Forecast horizon= 12 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-lift | 62.96% | 10 | 91.14% | 10 | 2.732 | 10 |
| ADL-own | 45.80% | 8 | 37.22% | 7 | 1.865 | 8 |
| ADL | 36.55% | 3 | 35.28% | 2 | 1.680 | 5 |
| ADL-own-EWC | 43.88% | 7 | 37.13% | 6 | 1.874 | 9 |
| ADL-own-IC | 47.38% | 9 | 36.51% | 4 | 1.852 | 7 |
| ADL-IC | 36.12% | 1 | 34.93% | 1 | 1.670 | 4 |
| ADL-EWC | 36.80% | 6 | 35.77% | 3 | 1.693 | 6 |
| ADL-DI | 36.64% | 4 | 37.40% | 9 | 1.506 | 2 |
| ADL-DI-IC | 36.49% | 2 | 37.00% | 5 | 1.487 | 1 |
| ADL-DI-EWC | 36.76% | 5 | 37.25% | 8 | 1.509 | 3 |

8.4 Forecasting performance regarding each product category

We explore the improvement by the IC method and the EWC method on the convention models regarding each product category. Table 6a shows the corresponding results of the ADL-own model, the ADL-own-EWC model, and the ADL-own-IC model. The figures are highlighted in yellow if the ADL-own-IC model and the ADL-own-EWC model outperform the ADL-own model. The results suggest that the ADL-own-EWC model outperform the ADL-own model for the majority of the product categories for the MAPE and the SMAPE, and has mixed results for the MASE. The ADL-own-IC model outperforms the ADL-own model for almost all the product categories for all the forecast horizons and error measures.

Table 6a. Forecasting performance for each product category: the ADL-own model, ADL-own-EWC model, and ADL-own-IC model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Rolling, Forecast horizon=1 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** |
| Beer | 22.16% | 22.49% | 21.30% | 22.46% | 22.81% | 21.65% | 0.788 | 0.800 | 0.759 |
| Carbonated beverages | 64.50% | 65.85% | 67.07% | 35.86% | 35.97% | 33.20% | 0.520 | 0.530 | 0.527 |
| Coffee | 30.56% | 30.49% | 29.36% | 29.78% | 30.20% | 29.26% | 0.603 | 0.606 | 0.588 |
| Frozen pizza | 27.16% | 26.81% | 26.56% | 29.65% | 29.42% | 28.56% | 1.038 | 1.033 | 1.028 |
| Household cleaner | 22.46% | 22.22% | 22.23% | 23.30% | 23.03% | 22.09% | 0.896 | 0.886 | 0.850 |
| Hotdog | 46.73% | 46.44% | 49.86% | 46.10% | 47.07% | 45.73% | 0.777 | 0.797 | 0.773 |
| Laundry detergent | 49.92% | 48.30% | 42.79% | 41.06% | 40.61% | 39.32% | 0.621 | 0.629 | 0.618 |
| Margarine/Butter | 20.79% | 19.98% | 20.01% | 21.39% | 20.57% | 20.21% | 0.513 | 0.498 | 0.477 |
| Mayonnaise | 22.76% | 22.63% | 22.55% | 20.52% | 20.46% | 20.99% | 0.769 | 0.770 | 0.793 |
| Mustard & ketchup | 31.82% | 32.35% | 29.48% | 27.34% | 27.97% | 26.19% | 0.582 | 0.590 | 0.533 |
| Peanut butter | 28.58% | 29.52% | 27.17% | 24.34% | 25.27% | 24.14% | 0.639 | 0.664 | 0.618 |
| Salty snacks | 31.84% | 30.27% | 27.90% | 26.58% | 26.03% | 25.85% | 0.666 | 0.700 | 0.639 |
| Soup | 28.36% | 29.61% | 27.50% | 25.52% | 26.16% | 25.70% | 0.329 | 0.335 | 0.330 |
| Sugar substitutes | 42.34% | 41.84% | 41.96% | 38.74% | 38.44% | 38.41% | 0.971 | 0.963 | 0.963 |
| Toothpaste | 45.81% | 45.43% | 46.55% | 47.12% | 47.17% | 47.31% | 0.436 | 0.432 | 0.428 |
| Rolling, Forecast horizon=4 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** |
| Beer | 22.47% | 22.46% | 21.45% | 22.88% | 22.89% | 21.92% | 0.813 | 0.813 | 0.783 |
| Carbonated beverages | 44.84% | 43.87% | 48.04% | 37.23% | 36.82% | 34.61% | 0.541 | 0.545 | 0.544 |
| Coffee | 30.67% | 30.55% | 29.61% | 30.38% | 30.85% | 29.92% | 0.632 | 0.631 | 0.614 |
| Frozen pizza | 28.30% | 28.14% | 27.89% | 30.62% | 30.51% | 29.75% | 0.665 | 0.666 | 0.656 |
| Household cleaner | 23.68% | 23.29% | 23.04% | 24.65% | 24.15% | 23.07% | 0.936 | 0.917 | 0.878 |
| Hotdog | 43.96% | 43.52% | 46.33% | 45.16% | 46.09% | 44.76% | 0.780 | 0.796 | 0.776 |
| Laundry detergent | 53.08% | 50.29% | 44.44% | 43.27% | 42.47% | 40.49% | 0.626 | 0.629 | 0.619 |
| Margarine/Butter | 20.74% | 20.24% | 19.51% | 22.05% | 21.52% | 20.34% | 0.546 | 0.538 | 0.496 |
| Mayonnaise | 25.59% | 25.62% | 24.78% | 22.34% | 22.42% | 22.32% | 0.851 | 0.862 | 0.860 |
| Mustard & ketchup | 38.02% | 38.93% | 33.88% | 30.57% | 31.55% | 27.88% | 0.679 | 0.693 | 0.581 |
| Peanut butter | 33.05% | 33.88% | 29.95% | 27.11% | 27.59% | 25.64% | 0.740 | 0.761 | 0.675 |
| Salty snacks | 33.89% | 31.41% | 28.71% | 27.02% | 26.12% | 25.39% | 0.663 | 0.689 | 0.630 |
| Soup | 30.32% | 31.88% | 28.98% | 26.71% | 27.26% | 26.59% | 0.343 | 0.348 | 0.345 |
| Sugar substitutes | 42.93% | 42.74% | 42.81% | 39.37% | 39.25% | 39.07% | 0.997 | 0.995 | 0.993 |
| Toothpaste | 46.80% | 45.65% | 45.01% | 47.30% | 46.94% | 45.18% | 0.444 | 0.442 | 0.431 |
| Rolling, Forecast horizon=12 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** | **ADL-own** | **ADL-own-EWC** | **ADL-own-IC** |
| Beer | 21.63% | 21.59% | 20.58% | 22.28% | 22.18% | 21.26% | 0.784 | 0.781 | 0.757 |
| Carbonated beverages | 50.23% | 46.62% | 55.14% | 41.56% | 40.78% | 38.41% | 0.625 | 0.617 | 0.625 |
| Coffee | 31.28% | 31.00% | 30.10% | 31.59% | 31.94% | 31.05% | 0.707 | 0.703 | 0.683 |
| Frozen pizza | 32.09% | 31.84% | 31.10% | 31.91% | 31.64% | 30.43% | 0.816 | 0.813 | 0.798 |
| Household cleaner | 32.83% | 32.37% | 32.80% | 27.02% | 26.19% | 25.36% | 0.974 | 0.945 | 0.916 |
| Hotdog | 42.46% | 42.84% | 44.59% | 45.04% | 46.15% | 44.66% | 0.768 | 0.784 | 0.764 |
| Laundry detergent | 54.30% | 50.57% | 45.40% | 44.10% | 43.28% | 41.45% | 0.592 | 0.589 | 0.587 |
| Margarine/Butter | 22.29% | 22.07% | 20.90% | 22.68% | 22.43% | 20.77% | 0.537 | 0.531 | 0.483 |
| Mayonnaise | 27.32% | 27.38% | 25.61% | 23.49% | 23.57% | 22.70% | 0.905 | 0.913 | 0.882 |
| Mustard & ketchup | 40.49% | 42.76% | 35.96% | 31.95% | 33.28% | 28.90% | 0.758 | 0.792 | 0.635 |
| Peanut butter | 42.00% | 42.70% | 38.06% | 31.96% | 32.20% | 30.13% | 0.924 | 0.946 | 0.834 |
| Salty snacks | 35.76% | 31.62% | 29.32% | 28.21% | 26.35% | 25.31% | 0.752 | 0.680 | 0.697 |
| Soup | 35.69% | 37.22% | 33.97% | 29.94% | 30.48% | 29.74% | 0.432 | 0.439 | 0.432 |
| Sugar substitutes | 42.05% | 41.84% | 42.49% | 38.90% | 38.68% | 38.66% | 0.955 | 0.950 | 0.952 |
| Toothpaste | 49.51% | 48.10% | 47.22% | 48.86% | 48.06% | 45.77% | 0.463 | 0.459 | 0.448 |

Table 6b shows the corresponding results of the ADL model, the ADL-EWC model, and the ADL-IC model. The figures are highlighted in yellow if the ADL-EWC model and the ADL-IC model outperform the ADL-own model. The results suggest that the ADL-EWC model outperforms the ADL model for a number of product categories. The ADL-IC model outperforms the ADL model for almost all the product categories for all the forecast horizons and error measures.

Table 6b. Forecasting performance for each product category: the ADL model, ADL-EWC model, and ADL-IC model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Rolling, Forecast horizon=1 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL** | **ADL-EWC** | **ADL-IC** | **ADL** | **ADL-EWC** | **ADL-IC** | **ADL** | **ADL-EWC** | **ADL-IC** |
| Beer | 21.78% | 21.85% | 21.20% | 21.60% | 21.79% | 21.20% | 0.752 | 0.758 | 0.739 |
| Carbonated beverages | 27.68% | 28.36% | 28.49% | 26.92% | 27.43% | 25.93% | 0.467 | 0.488 | 0.481 |
| Coffee | 30.07% | 30.72% | 29.53% | 29.87% | 30.91% | 29.46% | 0.599 | 0.610 | 0.585 |
| Frozen pizza | 26.43% | 26.01% | 25.58% | 28.64% | 28.24% | 27.38% | 1.015 | 1.009 | 1.006 |
| Household cleaner | 22.30% | 21.97% | 22.02% | 23.01% | 22.68% | 21.71% | 0.888 | 0.875 | 0.839 |
| Hotdog | 44.96% | 45.01% | 47.80% | 44.52% | 45.80% | 44.00% | 0.764 | 0.801 | 0.762 |
| Laundry detergent | 46.86% | 46.14% | 39.83% | 39.93% | 39.84% | 37.36% | 0.567 | 0.586 | 0.565 |
| Margarine/Butter | 20.60% | 19.98% | 20.27% | 20.97% | 20.39% | 20.49% | 0.499 | 0.490 | 0.488 |
| Mayonnaise | 23.99% | 24.24% | 23.56% | 21.53% | 21.84% | 21.77% | 0.806 | 0.825 | 0.823 |
| Mustard & ketchup | 30.21% | 30.57% | 29.14% | 26.32% | 26.59% | 26.52% | 0.554 | 0.563 | 0.549 |
| Peanut butter | 26.37% | 26.41% | 25.60% | 22.76% | 23.07% | 23.12% | 0.592 | 0.604 | 0.582 |
| Salty snacks | 28.41% | 28.38% | 25.06% | 24.87% | 25.15% | 23.85% | 0.581 | 0.651 | 0.566 |
| Soup | 27.57% | 28.04% | 26.56% | 25.42% | 25.84% | 25.38% | 0.335 | 0.336 | 0.335 |
| Sugar substitutes | 42.39% | 41.94% | 41.98% | 38.81% | 38.57% | 38.44% | 0.972 | 0.965 | 0.963 |
| Toothpaste | 46.31% | 47.18% | 47.26% | 46.48% | 46.53% | 47.00% | 0.467 | 0.466 | 0.466 |
| Rolling, Forecast horizon=4 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL** | **ADL-EWC** | **ADL-IC** | **ADL** | **ADL-EWC** | **ADL-IC** | **ADL** | **ADL-EWC** | **ADL-IC** |
| Beer | 21.86% | 21.58% | 21.24% | 21.92% | 21.77% | 21.44% | 0.775 | 0.772 | 0.762 |
| Carbonated beverages | 26.10% | 26.51% | 27.14% | 26.93% | 27.09% | 26.16% | 0.495 | 0.504 | 0.513 |
| Coffee | 30.22% | 30.83% | 29.88% | 30.44% | 31.49% | 30.16% | 0.628 | 0.635 | 0.614 |
| Frozen pizza | 26.93% | 26.56% | 26.36% | 28.68% | 28.30% | 27.69% | 0.629 | 0.628 | 0.624 |
| Household cleaner | 23.46% | 22.97% | 22.86% | 24.31% | 23.75% | 22.72% | 0.927 | 0.906 | 0.869 |
| Hotdog | 42.32% | 42.37% | 43.96% | 43.70% | 45.19% | 42.80% | 0.769 | 0.798 | 0.767 |
| Laundry detergent | 50.73% | 48.36% | 41.81% | 42.62% | 41.46% | 38.74% | 0.597 | 0.602 | 0.583 |
| Margarine/Butter | 20.72% | 20.35% | 19.94% | 21.75% | 21.46% | 20.79% | 0.536 | 0.532 | 0.510 |
| Mayonnaise | 26.59% | 26.81% | 25.70% | 22.92% | 23.29% | 22.86% | 0.880 | 0.896 | 0.887 |
| Mustard & ketchup | 33.86% | 33.80% | 31.95% | 27.89% | 27.73% | 27.37% | 0.586 | 0.589 | 0.569 |
| Peanut butter | 28.51% | 28.41% | 26.72% | 24.32% | 24.49% | 23.87% | 0.664 | 0.675 | 0.619 |
| Salty snacks | 30.49% | 30.11% | 26.10% | 25.87% | 26.14% | 24.15% | 0.628 | 0.674 | 0.605 |
| Soup | 29.07% | 29.31% | 27.59% | 26.73% | 26.96% | 26.36% | 0.364 | 0.363 | 0.364 |
| Sugar substitutes | 42.98% | 42.82% | 42.86% | 39.44% | 39.35% | 39.13% | 0.998 | 0.996 | 0.993 |
| Toothpaste | 47.32% | 46.62% | 45.77% | 47.09% | 46.70% | 45.18% | 0.473 | 0.469 | 0.459 |
| Rolling, Forecast horizon=12 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL** | **ADL-EWC** | **ADL-IC** | **ADL** | **ADL-EWC** | **ADL-IC** | **ADL** | **ADL-EWC** | **ADL-IC** |
| Beer | 20.72% | 20.31% | 20.20% | 21.12% | 20.77% | 20.73% | 0.742 | 0.732 | 0.735 |
| Carbonated beverages | 29.55% | 29.58% | 30.42% | 28.69% | 28.60% | 27.43% | 0.678 | 0.700 | 0.704 |
| Coffee | 30.90% | 31.22% | 30.32% | 31.39% | 32.18% | 30.99% | 0.701 | 0.702 | 0.682 |
| Frozen pizza | 30.17% | 29.87% | 29.02% | 30.32% | 29.95% | 28.77% | 0.674 | 0.672 | 0.663 |
| Household cleaner | 32.59% | 32.14% | 32.53% | 26.71% | 25.91% | 24.94% | 0.966 | 0.937 | 0.905 |
| Hotdog | 41.16% | 41.75% | 42.72% | 43.72% | 45.35% | 42.97% | 0.760 | 0.781 | 0.762 |
| Laundry detergent | 52.39% | 49.03% | 42.73% | 43.84% | 42.65% | 39.81% | 0.577 | 0.575 | 0.557 |
| Margarine/Butter | 22.01% | 21.91% | 21.33% | 22.08% | 22.03% | 21.16% | 0.517 | 0.515 | 0.493 |
| Mayonnaise | 27.83% | 28.17% | 26.42% | 23.47% | 23.91% | 22.89% | 0.910 | 0.929 | 0.898 |
| Mustard & ketchup | 35.44% | 36.05% | 33.39% | 29.19% | 29.37% | 28.51% | 0.638 | 0.648 | 0.616 |
| Peanut butter | 33.09% | 33.36% | 31.16% | 27.54% | 27.85% | 27.18% | 0.791 | 0.809 | 0.737 |
| Salty snacks | 32.13% | 31.28% | 27.30% | 28.06% | 27.95% | 25.82% | 0.724 | 0.745 | 0.699 |
| Soup | 31.15% | 31.62% | 29.58% | 28.42% | 28.85% | 28.14% | 0.428 | 0.430 | 0.428 |
| Sugar substitutes | 42.10% | 41.88% | 42.55% | 38.95% | 38.74% | 38.70% | 0.955 | 0.951 | 0.953 |
| Toothpaste | 49.08% | 48.05% | 47.08% | 48.91% | 48.23% | 46.19% | 0.462 | 0.458 | 0.444 |

Table 6c shows the corresponding results of the ADL-DI model, the ADL-DI-EWC model, and the ADL-DI-IC model. The figures are again highlighted in yellow if the ADL-DI-EWC model and the ADL-DI-IC model outperform the ADL-DI model. The results suggest that the ADL DI-EWC model outperforms the ADL-DI model for the majority of the product categories. The ADL-DI-IC model outperforms the ADL-DI model for almost all the product categories for all the forecast horizons and error measures.

Table 6c. Forecasting performance for each product category: the ADL-DI model, ADL-EWC-DI model, and ADL-IC-DI model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Rolling, Forecast horizon=1 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** |
| Beer | 21.47% | 21.22% | 21.40% | 21.22% | 21.05% | 21.33% | 0.733 | 0.733 | 0.738 |
| Carbonated beverages | 53.90% | 58.92% | 52.51% | 26.01% | 25.79% | 26.29% | 0.424 | 0.432 | 0.423 |
| Coffee | 30.49% | 30.23% | 30.03% | 29.17% | 29.46% | 29.49% | 0.592 | 0.596 | 0.596 |
| Frozen pizza | 26.60% | 25.79% | 26.05% | 28.04% | 27.40% | 26.96% | 1.035 | 1.025 | 1.024 |
| Household cleaner | 22.02% | 21.96% | 21.26% | 22.64% | 22.53% | 21.36% | 0.876 | 0.871 | 0.826 |
| Hotdog | 47.69% | 47.51% | 49.84% | 47.50% | 48.21% | 46.31% | 0.803 | 0.814 | 0.770 |
| Laundry detergent | 37.16% | 38.52% | 35.66% | 36.63% | 36.22% | 36.39% | 0.564 | 0.574 | 0.557 |
| Margarine/Butter | 19.53% | 19.45% | 19.26% | 20.24% | 20.00% | 19.37% | 0.467 | 0.458 | 0.438 |
| Mayonnaise | 23.16% | 23.73% | 22.86% | 20.93% | 21.69% | 20.97% | 0.793 | 0.824 | 0.796 |
| Mustard & ketchup | 30.84% | 30.35% | 30.40% | 28.42% | 27.87% | 28.99% | 0.571 | 0.563 | 0.578 |
| Peanut butter | 24.98% | 24.31% | 24.29% | 21.93% | 21.82% | 22.06% | 0.566 | 0.561 | 0.562 |
| Salty snacks | 25.97% | 26.15% | 25.31% | 24.60% | 25.22% | 24.81% | 0.596 | 0.678 | 0.598 |
| Soup | 28.00% | 29.09% | 27.59% | 25.50% | 26.14% | 25.93% | 0.324 | 0.326 | 0.326 |
| Sugar substitutes | 45.29% | 45.73% | 44.99% | 47.94% | 48.55% | 47.86% | 1.007 | 1.020 | 1.004 |
| Toothpaste | 44.79% | 46.42% | 45.05% | 48.36% | 48.75% | 46.46% | 0.454 | 0.461 | 0.448 |
| Rolling, Forecast horizon=4 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** |
| Beer | 21.64% | 21.33% | 21.64% | 21.43% | 21.22% | 21.60% | 0.754 | 0.753 | 0.760 |
| Carbonated beverages | 31.65% | 32.57% | 31.84% | 25.27% | 24.99% | 25.21% | 0.414 | 0.422 | 0.412 |
| Coffee | 29.43% | 29.20% | 29.22% | 28.47% | 28.84% | 29.00% | 0.596 | 0.597 | 0.603 |
| Frozen pizza | 27.04% | 26.46% | 26.33% | 27.72% | 27.16% | 26.49% | 0.650 | 0.642 | 0.638 |
| Household cleaner | 23.62% | 23.39% | 22.87% | 24.42% | 24.06% | 23.11% | 0.932 | 0.918 | 0.881 |
| Hotdog | 45.57% | 45.40% | 47.19% | 47.07% | 47.88% | 45.78% | 0.811 | 0.823 | 0.782 |
| Laundry detergent | 37.24% | 37.51% | 35.62% | 37.25% | 36.44% | 36.97% | 0.567 | 0.566 | 0.565 |
| Margarine/Butter | 19.70% | 19.74% | 19.36% | 20.90% | 20.75% | 19.90% | 0.493 | 0.490 | 0.457 |
| Mayonnaise | 25.27% | 26.08% | 24.74% | 22.19% | 23.05% | 22.04% | 0.851 | 0.892 | 0.848 |
| Mustard & ketchup | 32.92% | 32.40% | 32.93% | 29.59% | 28.99% | 30.39% | 0.604 | 0.592 | 0.625 |
| Peanut butter | 26.45% | 25.77% | 25.18% | 22.83% | 22.78% | 22.47% | 0.601 | 0.594 | 0.579 |
| Salty snacks | 27.94% | 27.45% | 26.66% | 26.66% | 26.76% | 26.46% | 0.643 | 0.712 | 0.636 |
| Soup | 29.37% | 30.74% | 28.52% | 26.42% | 26.96% | 26.48% | 0.341 | 0.339 | 0.344 |
| Sugar substitutes | 44.21% | 44.55% | 43.97% | 44.30% | 44.80% | 44.09% | 0.968 | 0.980 | 0.964 |
| Toothpaste | 44.77% | 45.49% | 44.62% | 47.81% | 48.02% | 45.69% | 0.434 | 0.441 | 0.429 |
| Rolling, Forecast horizon=12 | MAPE | | | SMAPE | | | MASE | | |
| **category** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** | **ADL-DI** | **ADL-DI-EWC** | **ADL-DI-IC** |
| Beer | 20.21% | 20.06% | 20.16% | 20.25% | 20.06% | 20.35% | 0.718 | 0.714 | 0.722 |
| Carbonated beverages | 33.17% | 32.99% | 33.59% | 28.85% | 28.41% | 28.35% | 0.525 | 0.538 | 0.517 |
| Coffee | 28.81% | 28.63% | 28.52% | 28.60% | 28.97% | 29.11% | 0.634 | 0.634 | 0.636 |
| Frozen pizza | 29.61% | 29.17% | 28.66% | 29.38% | 28.95% | 27.88% | 0.607 | 0.601 | 0.586 |
| Household cleaner | 33.85% | 33.84% | 34.19% | 27.27% | 26.73% | 26.09% | 0.985 | 0.966 | 0.942 |
| Hotdog | 43.50% | 44.12% | 44.78% | 47.26% | 48.36% | 45.93% | 0.787 | 0.800 | 0.755 |
| Laundry detergent | 37.32% | 37.30% | 35.76% | 37.66% | 37.09% | 37.55% | 0.536 | 0.537 | 0.535 |
| Margarine/Butter | 20.53% | 20.55% | 20.20% | 21.11% | 20.97% | 20.19% | 0.477 | 0.475 | 0.449 |
| Mayonnaise | 25.60% | 26.06% | 25.25% | 22.53% | 23.11% | 22.53% | 0.869 | 0.899 | 0.871 |
| Mustard & ketchup | 33.98% | 33.86% | 35.49% | 29.70% | 29.57% | 31.46% | 0.669 | 0.668 | 0.716 |
| Peanut butter | 29.88% | 29.22% | 28.71% | 25.49% | 25.43% | 25.24% | 0.687 | 0.679 | 0.663 |
| Salty snacks | 31.23% | 30.15% | 29.46% | 29.89% | 29.76% | 29.43% | 0.805 | 0.841 | 0.789 |
| Soup | 32.99% | 34.67% | 31.67% | 28.76% | 29.71% | 28.62% | 0.411 | 0.414 | 0.411 |
| Sugar substitutes | 41.69% | 41.88% | 41.41% | 43.55% | 43.82% | 43.25% | 0.916 | 0.921 | 0.910 |
| Toothpaste | 44.59% | 44.80% | 44.79% | 48.10% | 47.96% | 46.20% | 0.417 | 0.418 | 0.412 |

1. **Conclusion and Future research**

Grocery retailers have been struggling with producing accurate sales forecasts to effectively manage their inventory planning and customer satisfaction. In practice, many retailers use simple univariate models with adjustments for incoming promotional events. Some recent studies focused on taking advantage of the impact of promotional activities. For example, Gur Ali et al. (2009) proposed models with sophisticated function forms (e.g., the regression tree model) with the price and promotional information of the focal product. Huang et al. (2014) incorporated the competitive promotional information within the same product category by resorting to variable selection methods and the principle component analysis which mitigated the problem of high dimensionality. Ma et al. (2016) integrated the promotional information both within the same product category and across difficult product categories.

These studies all assume constant effectiveness of the promotional activities. However, evidence show that the effectiveness of the promotional activities may change over time because of the impact of many influencing factors including the change of economic condition, the change of the consumer taste, and new competition entry etc. These factors may not be observable (even we know there is a new competitor entering the market, we may not be able to immediately collect the information and incorporate them into the model. Unless the unobserved variables of these factors are orthogonal to all the price and promotional variables which are already included in the model, the effectiveness of these variables will change. The model will be subject to structural break, and produce biased and

less accurate forecasts.

In this study, we propose more effective forecasting method by taking into account the unobservable change in the effectiveness of the promotional activities and the associated issue of structural break. We employ two different methods including the intercept correction and the estimation window combining method discriminately with the general-to-specific ADL model. The IC method offset the potential forecast bias by adding the estimate of the bias back to the forecasts at a cost of inflated forecast error variance. The EWC method combines the sets of forecasts by the same model with different estimation windows and tries to achieve a trade-off between the forecast bias and the forecast error variance. The two methods have been proved successful in macroeconomic forecasting. Our study is the first study focus on the unobserved changing effectiveness of promotional activities on a forecasting perspective. Our study is also the first study which investigates if we can generate more accurate forecasts with these two methods based on the general-to-specific ADL model. We also conduct our evaluation for the model which exclusively has the promotional information of the focal product and for the model with competitive promotional information within the same product category, and our solution is useful for both manufacturers and retailers.

In this study, we evaluate the performance of various candidate models in forecasting retailer product sales at the SKU level for 128 SKUs across 15 product categories. We have the following findings:

1. The ADL-own-IC model, the ADL-IC model, and the ADL-DI-IC model outperform the ADL-own model, the ADL model, and the ADL-DI model respectively.
2. The ADL-IC model and the ADL-DI-IC model generally produces the most accurate forecasts.
3. The ADL-own-IC model outperforms the base-time-lift method and the ADL-own model. The model is especially valuable to manufacturers when they cannot get access to competitive information the retailers.
4. The improvement by implementing the IC method mainly come from the period when the focal product is not being promoted.
5. The EWC method only moderately improve the forecasting performance of the ADL-own model, the ADL model, and the ADL-DI model.

Overall, we recommend the ADL-IC model and the ADL-DI-IC model to forecast retailer product sales at the SKU level and we also recommend using the ADL-own-IC model for manufacturers when competitive promotional information is not an option. Table 7 shows the percentage reductions regarding various forecast horizons and error measures by the IC method and the EWC method. For example, we can reduce the MAPE by 4.78% if we adopt the ADL-own-IC model compared to the ADL-own model for one to twelve weeks ahead forecast horizon. We can reduce the SMAPE by 3.97% if we adopt the ADL-IC model compared to the ADL model for one to four weeks ahead forecast horizon.

Table 7. Forecasting performance regarding percentage reductions in various error measures

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Models | Forecast horizon= 1 | | | Forecast horizon= 4 | | | Forecast horizon= 12 | | |
| MAPE | SMAPE | MASE | MAPE | SMAPE | MASE | MAPE | SMAPE | MASE |
| ADL-own | compared to ADL-own | | | compared to ADL-own | | | compared to ADL-own | | |
| ADL-own-IC | 2.59% | 2.44% | 2.20% | 4.62% | 4.17% | 3.57% | 4.78% | 5.03% | 4.45% |
| ADL-own-EWC | 0.32% | -0.30% | -0.86% | 1.24% | 0.10% | -0.66% | 1.88% | 0.53% | 0.55% |
| ADL | compared to ADL | | | compared to ADL | | | compared to ADL | | |
| ADL-IC | 2.61% | 1.88% | 1.12% | 4.53% | 3.45% | 2.14% | 4.76% | 3.97% | 2.39% |
| ADL-EWC | -0.22% | -0.74% | -1.85% | 0.80% | -0.15% | -0.93% | 0.87% | 0.01% | -0.59% |
| ADL-DI | compared to ADL-DI | | | compared to ADL-DI | | | compared to ADL-DI | | |
| ADL-DI-IC | 1.11% | 1.08% | 1.30% | 1.33% | 1.55% | 1.48% | 0.95% | 1.43% | 1.40% |
| ADL-DI-EWC | -1.68% | -0.37% | -1.44% | -0.32% | -0.10% | -1.13% | -0.09% | -0.11% | -0.64% |

There are some studies including [Foekens, Leeflang et al. (1999)](#_ENREF_28) which modelled the effectiveness of the price with the previous price of the product and for the promotion with the recency and frequency of previous promotion. The model is however used to estimate and interpret the presume changing process of the effectiveness of the price and promotion and was not developed for forecasting purposes. We have conducted preliminary evaluation for the ADL model with autoregressive functions for the effectiveness of the price and promotion. The model however generated poor forecasts. One possible reason may be that it is difficult to model appropriately the changing process of the price and promotions and it is very easy to make the model very too complex by engaging with time-varying parameters. The model suffers from both low robustness and poor parsimony which are critically important regarding forecasting accuracy.

There are also methods alternative to the IC method and the EWC method which also take into account structural break and forecast bias. [Castle, Doornik et al. (2008)](#_ENREF_11) and [Hendry and Krolzig (2001)](#_ENREF_32) proposed to construct the ADL model with dummy variables for each of the observations and then recursively simplify the model using the *Autometrics* algorithm (what it basically does is to simplify the model block by block pretending there are only a section of the dummy variable existing and eventually combine the dummy variables which are retained in previous steps, and so forth). The final model will not be subject to structural break and thus would ideally generate unbiased forecasts. However, the method comes with the cost of losing information (e.g. the information contained in the observations offset by the retained dummy variables) and also may not always reduce the out-of-sample forecast bias (e.g., when there is one structural break close to the forecast origin within the estimation period). The performance of the method is an empirical question for the product sales forecasting at the SKU level. We leave the evaluation of this method to the next stage of our research.

There are some limitations for this study. We used the sequential Chow test to detect the existence of structural break. We may adopt those tests which allow for heteroskedasticity and multiple structural breaks ([Andrews 1993](#_ENREF_4), [Andrews and Ploberger 1994](#_ENREF_5), [Bai and Perron 2003](#_ENREF_9)). It is possible to see different performance by the IC method and the EWC method with these tests compared to the ad hoc sequential Chow test we use in this study. Also, there could be potentials to get better forecasting performance with variants of the IC method and the EWC method. In this study, we add the estimated bias directly to the forecasts. Clements and Hendry (1999) summarized other correction schemes which have different intrinsic characteristics regarding their impact on bias correction and forecast error variance inflation. In this study, we recommend the ADL-own-IC model (for manufacturers) and the ADL-IC model (for retailers). Ma et al. (2016) proposed models which further integrate both the intra and the inter category promotional information. It would be interesting to investigate if the model with additional promotional information would still be subject to structural break and generate biased forecasts, and, if yes, whether we can further improve the forecasting performance by taking into account the change of the effectiveness of all these variables.

At last, we thank the IRI company to make the data available for the evaluation of our models.

**Reference:**

Aburto, L. and R. Weber (2007). "Improved supply chain management based on hybrid demand forecasts." Applied Soft Computing **7**: 136-144.

Ali, M. and J. Boylan (2011). "Feasibility principles for Downstream Demand Inference in supply chains." Journal of the Operational Research Society **62**.

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.

Andrews, R. L., et al. (2008). "Estimating the SCAN\*PRO model of store sales: HB, FM or just OLS?" international Journal of research in marketing **25**(1): 22-33.

Arenas, T., et al. (2013). "Analysis of judgmental adjustments in the presence of promotions." International Journal of Forecasting **29**(2).

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

Bai, J. and P. Perron (2003). "Computation and Analysis of Multiple Structural-Change Models." Journal of Applied Econometrics **18**: 1-22.

Bronnenberg, B. J., et al. (2008). "The IRI Marketing Data Set." Marketing Science **27**(4): pp. 745–748.

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

Clark, T. E. and M. W. McCracken (2007). Forecasting with Small Macroeconomic VARs in the Presence of Instabilities. Finance and Economics Discussion Series

Divisions of Research & Statistics and Monetary Affairs

Federal Reserve Board, Washington, D.C.

Clements, M. and D. Hendry (1998). Forecasting Economic Time Series, Cambridge University Press.

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.

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).

Cooper, L. G., et al. (1999). "Promocast": a New Forecasting Method for Promotion Planning." Marketing Science **18**(3): 301-316.

Cooper, L. G. and G. Giuffrida (2000). "Turning Datamining into a Management Science Tool: New Algorithms and Empirical Results." Management Science **46**(2): 249.

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).

Davydenko, A. and R. Fildes (2013). "Measuring forecasting accuracy: the case of judgmental adjustments to SKU-level demand forecasts." International Journal of Forecasting **29**(3).

Dekimpe, M., et al. (1999). "Long-run effects of price promotions in scanner markets." Journal of Econometrics **89**(1/2): 261-291.

Dekker, M., et al. (2004). "How to use aggregation and combined forecasting to improve seasonal demand forecasts." International Journal of Production Economics **90**(2): 151-167.

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

Fildes, R., et al. (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., et al. (2008). "Forecasting and operational research: A review." Journal of the Operational Research Society **59**.

Fildes, R. and H. Stekler (2002). "The state of macroeconomic forecasting." Journal of Macroeconomics **24**(4): 435-468.

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.

Goodwin, P. (2002). "Integrating management judgment and statistical methods to improve short-term forecasts." Omega **30**(2): 127-135.

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.

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.

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.

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).

Loeb, W. (2015). "Unrelenting Competition: The Biggest Retail Story of 2015." 2016.

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.

Mace, S. and S. A. Neslin (2004). "The determinants of pre- and postpromotion dips in sales of frequently purchased goods." Journal of Marketing Research **XLI**: 339-350.

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.

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.dynamicaction.com/WS-2015-06-IHL-Ghost-Economy-Haunting-of-Returns-AR_LP.html>.

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.

Pesaran, M. H. and A. Timmermann (2005). "Small sample properties of forecasts from autoregressive models under structural breaks." Journal of econometrics **129**(1-2): 183-217.

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.

Tashman, L. J. (2000). "Out-of-sample tests of forecasting accuracy: an analysis and review " International Journal of Forecasting **16**(4).

Trusov, M., et al. (2006). "Retailer Promotion Planning: Improving Forecasting Accuracy And Interpretability." Journal of Interactive Marketing **20**(3-4): 71-81.

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).

Wittink, D., et al. (1988). SCAN\*PRO: the estimation, validation and use of promotional effects based on scanner data. Internal paper, Cornell University.

1. Strictly speaking, the forecast bias comes from the change of the deterministic mean of the model due to the change of the model parameters. However, there is a very rare possibility that the deterministic mean could retain unchanged even if the parameters all change (but in a very specific way). Under such a circumstance, there will be no forecast bias even when the model is subject to structural break. However, in this study we do not consider this situation because it only happens theoretically when very restrictive conditions are met. Details based on an example of a VAR model can be found in Clements and Hendry (1999). Thereafter in this study, we assume structural breaks lead to forecast bias for the models. [↑](#footnote-ref-1)
2. This example demonstrates the issue of structural break using a simple static model. The analytical evidence for dynamic models can be found in Clements and Hendry (1999) and Pesaran and Timmerman (2005, 2007). [↑](#footnote-ref-2)
3. This setting is very common in the retailer context. In this example we artificially make up the data series but we keep the data series to be stationary. [↑](#footnote-ref-3)
4. We may have an alternative example where the sales generally decrease but also become more responsive to temporary price reductions. This may be caused by new product introduction, more competitive promotional activities by other products, or the change of economic conditions and consumer taste which are unobservable. [↑](#footnote-ref-4)
5. The Chow test is a variant of F-test which compares the fitting of the model before and after the structural break. It assumes the locations of the structural known a priori. [↑](#footnote-ref-5)
6. To mitigate the multiple comparison problem, we adopt very small threshold (e.g., 0.0001) for the p-value of the sequential test. [↑](#footnote-ref-6)
7. In the simulation, we assume there is only one explanatory variable. Thus 30 observations are more than enough to estimate the model. [↑](#footnote-ref-7)
8. 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-8)
9. We count feature and display as one type. i.e., feature/display event. [↑](#footnote-ref-9)
10. 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*. [↑](#footnote-ref-10)
11. *L* is initially set as two but more lags of the variables are added to the general model if the model does not pass the specification test. [↑](#footnote-ref-11)
12. A very small p-value (i.e. 0.005) is used for the sequential Chow test to mitigate the multiple testing problem [↑](#footnote-ref-12)
13. 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-13)
14. We refer the period when the focal product is not being promoted (feature or display) as the ‘non-promoted’ period. [↑](#footnote-ref-14)