**Forecasting Product Sales for retailer product sales with the impact by the unobserved factors**

Taking into account the unobserved change of the effectiveness of the price and promotional information.

Tao Huang, Robert Fildes, Didier Soopramanien

Abstract

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

**Be careful: self- plagiarism when copy some sentence from previous papers**

Key words:

Sales Forecasting, Marketing analytics, Promotion

**Section 1: Introduction**

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

Retailers have been struggling with the situations of out-of-stock and over-stock for years. When a product is out-of-stock, retailers not only lose profits but also may lose the customers forever. Previous studies show that customers whom were once believed to either purchase alternative products or postpone their purchases when their preferred products are out of stock are actually more likely to switch to other stores and never come back ([Corsten and Gruen 2003](#_ENREF_15)). In practice, retailers try may deliberately increasing the inventory level (i.e. to over-stock) to avoid the out-of-stock condition, which however significantly raises inventory costs and reduces profits (Cooper, Baron et al. 1999). Under such a circumstance, retailers need to balance the loss due to running out-of-stock and the cost of higher inventory level. One of the keys to resolve this cost and service level dilemma is to generate accurate forecasts for the product sales (Corsten and Gruen 2003).

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

In practice, many retailers have been using a two stage ‘base-times-lift’ approach to generate forecasts for product sales at the SKU level. For example, retailers may generate a baseline forecast using simple exponential smoothing methods and then make adjustments for any incoming promotional event. The adjustments are usually made by brand/category managers. In the literature, a stream of studies have been devoted to help managers improve their adjustment procedure (Fildes et al., 2008; Goodwin, 2002; Lee et al., 2007; Nikolopoulos, 2010). Alternatively, some studies proposed model-based forecasting system estimate the adjustment (Cooper et al., 1999). Other studies including Kuo (2001), Aburto and Weber (2007) and Gur Ali et al. (2009) proposed machine learning algorithms which include the promotional information of the focal product. ([Huang, Fildes et al. 2014](#_ENREF_26)) is the first study which directly incorporates the promotional information of not only the focal product but also of the competitive products within the same product category. ([Ma, Fildes et al. 2016](#_ENREF_38)) further included the promotional information cross categories though using models with different structures and specification strategies.

The issue of structural breaks and forecast bias.

All these studies assume that the effect of the promotional activities does not change over time. In practice, however, this may not be true due to the impact of many influencing factors including the change of economic conditions, new legislation, the change of consumer tastes, and media habits, new competitor entry etc. ([Wildt 1976](#_ENREF_57), [Wildt and Winer 1983](#_ENREF_58)). Under such a circumstance, the model may potentially be subject to structural breaks which is defined as large change in the model with respect to the constant term or/and the parameter coefficients ([Armstrong 2001](#_ENREF_5)). As a result, the model may potentially produce biased and less accurate forecasts. The issue of structural breaks and forecast bias have been intensively address in the economics literature. For example, Pesaram and Timmerman xxx

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

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

**Literature review**

2.1 Existing studies forecasting retailer product sales

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

Some recent studies proposed more sophisticated models to directly forecast the product sales (other than adding the adjustment to the baseline forecast) and also take into account the promotional information of the focal product. Kuo (2001) and Aburto and Weber (2007) evaluated the performance of neural network algorithms in forecasting supermarket food products. [Gür Ali, SayIn et al. (2009)](#_ENREF_21) proposed support vector machine methods and regression tree methods to forecast retailer product sales at SKU level. Divakar et al. (2005) built the CHAN4CAST system which employed models of the regression form with information including past sales, trend, prices and promotions of the focal brand and the major competitors, and seasonality etc. However, the forecasting system were built at the brand/company level and only focused on the interaction between two brands (e.g., Coke and Pepsi). The situation for a retailer becomes very different as there can be hundreds of items competing with each other in a typical product category at any time (Cooper et al. 1999). Huang, Fildes and Soopramanien (2014) is the first study in the forecasting literature which proposed effective forecasting methods for retailer product sales at the SKU level with a formal procedure to select, refine, and incorporate competitive prices and competitive promotions from hundreds of competing items within the same product category. They built generate-to-specific econometric models with the most valuable competitive promotional information either selected by a LASSO/stepwise procedure or condensed by a principle component analysis. The method generated substantially more accurate forecasts across a range of product categories. Ma et al. (2016) further integrated the promotional information not only from the same product category (e.g., intra- category information) but also from other categories (e.g., inter- category information). Their proposed method relies on the LASSO algorithm for both variable selection and model specification.

2.2 The problem of structural break.

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

The effect of promotional activities

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

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

The effect of promotional activities can change over time

Some studies tend to explore the presumed ‘constant’ effect of the promotional activities on the product sales (or consumer preferences) under specific circumstance (e.g., [Hoch, Kim et al. 1995](#_ENREF_24), [Bijmolt, Heerde et al. 2005](#_ENREF_8)). There are many studies which have been devoted into exploring the changing effects of marketing activities (e.g. [Little 1966](#_ENREF_37), [Morrison 1966](#_ENREF_43), [Myers and Nicosia 1970](#_ENREF_46), [Myers 1971](#_ENREF_45), [Houston and Weiss 1975](#_ENREF_25), [Monroe and Guiltinan 1975](#_ENREF_42), [Moinpour, McCullough et al. 1976](#_ENREF_41), [Wildt 1976](#_ENREF_57), [Wichern and Jones 1977](#_ENREF_56), [Winer 1979](#_ENREF_59), [Mahajan, Bretschneider et al. 1980](#_ENREF_39)). Early studies argues that the effectiveness of promotions may change because of economic condition, legislation, consumer tastes, media habits, competition, and advertising etc. ([Wildt 1976](#_ENREF_57), [Wildt and Winer 1983](#_ENREF_58)). It is generally known that the effects of the marketing mix variables will change with different stages of the product life cycle ([Mahajan, Bretschneider et al. 1980](#_ENREF_39)). For instance, marketing theory suggests that the elasticities for marketing instruments (e.g. advertising, price, service, product quality, and packaging) are the highest at the growth stage of the product and the lowest at the maturity stage of the product ([Kotler 1997](#_ENREF_32)). The introduction of new products (especially the store-owned brand) may decrease the promotional elasticity of the premium national brand and increase the promotional elasticity of the second tier national brand ([Nijs, Dekimpe et al. 2001](#_ENREF_47), [Van Heerde, Srinivasan et al. 2008](#_ENREF_54)).

The effectiveness of promotions can change due to many reasons, as suggested by previous studies.

Intensive promotions can reduce consumers’ reference price ([Lattin and Bucklin 1989](#_ENREF_33), [Lichtenstein and Bearden 1989](#_ENREF_36), [Kalwani, Yim et al. 1990](#_ENREF_29), [Kalwani and Yim 1992](#_ENREF_28), [Foekens, S.H. Leeflang et al. 1999](#_ENREF_19), [Kopalle, Mela et al. 1999](#_ENREF_31), [Levy, Grewal et al. 2004](#_ENREF_35)), which accordingly changes the effects of promotions. For example, consumers may find the promotions less attractive if the products are promoted more frequently than before. The introduction of a new distribution channel can change the market response structure ([Verhoef, Neslin et al. 2007](#_ENREF_55)). For example, consumers may collect information in the newly constructed channel and adjust their reference price accordingly. The introduction of a new loyalty program can change the market response structure ([Leenheer, van Heerde et al. 2007](#_ENREF_34)). For example, retailers may launch promotional events to attract consumers from their competitors. However, when consumers become loyalty program members of a specific retailer, they receive saving rewards and direct discounts, and may find the promotions in other retailers less attractive. Accordingly, the termination of the existing loyalty program also changes the market response structure ([Melnyk and Bijmolt 2007](#_ENREF_40)). The relationship between product sales and the marketing mix variables may also change over time due to the evolving market structure specific to the retailer sales at the UPC level. As was discussed in section 3.2, the UPC set in the product categories may change considerably because ofthe introduction of new product UPCs or termination of existing product UPCs and changes in the assortment policies by the retailer (e.g. the retailer may decide to increase or reduce the number of UPCs in the product category) ([Bell, Bonfrer et al. 2005](#_ENREF_7)). That is, a single product UPC is competing with different sets of competitive products as time goes by, and the effects of the promotions on the focal product may change accordingly.

The change of the effectiveness of promotions has been applied to allocate budget.

Foekens, S.H. Leeflang et al. ([1999](#_ENREF_19)) extended the original SCAN\*PRO model to incorporate the time-varying effects of the marketing mix variables. In the extended model, the parameters of the marketing mix variables are functionally related to historical information of the focal brand and other competitive brands. For example, the intercept for the store and the price elasticity of the focal brand are related to previous price discounts of the focal brand and the competitive brands; the elasticities of the non-price promotions for the focal brand are related to the time since the most recent promotion for the focal brand and the competitive brands. The model aims to capture how the effects of the marketing mix variables change over time so that managers can allocate the marketing budget more efficiently. Kopalle, Mela et al. ([1999](#_ENREF_30)) also extended the SCAN\*PRO model in a similar manner to investigate the dynamic impact of promotions on the baseline sales. In their extended SCAN\*PRO model, the effects of price reductions are assumed to change according to previous discounting history. The results show that promotions increase the concurrent product sales but reduce the baseline sales.

However, only a few early studies have attempted to taken into account the change of the effect of marketing activities (e.g. advertising) over time in forecasting product sales ([Mahajan, Bretschneider et al. 1980](#_ENREF_39)). [Cooley and Prescott (1976)](#_ENREF_13) proposed models which allows the parameter to change in an autoregressive manner, say, , , where and are uncorrelated error terms. In an alternative form the parameters were modelled as a function of a constant term with a disturbance term, e.g., , where is the error term. The autoregressive variation model has been applied to capture how the effects of advertising change over time ([Pekelman and Edison 1980](#_ENREF_48)), but the random variation function was rarely used because it does not track the changing effects of the marketing mix variables over time (Wildt and Winer 1983).

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

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

When the effectiveness of the price and promotions on product sales change, 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_2)). 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 economics literature (e.g. [Cooper and Nelson 1975](#_ENREF_14), [Muellbauer 1994](#_ENREF_44), [Hendry 1995](#_ENREF_22), [Clements and Hendry 1999](#_ENREF_12), [Pesaran and Timmermann 2007](#_ENREF_49), [Castle, Doornik et al. 2008](#_ENREF_9)).

Previous studies showed analytical evidence for the impact of a structural break within the estimation sample on the model’s forecasting performance (e.g., [Pesaran and Timmermann, 2007)](#_ENREF_49). 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. We assume that the real demand follows the process below:

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*. 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 the most recent *m* observations . That is, the data at week *T*-*m+1* is the first observation in the estimation window. We have the OLS estimate as follows:

where and are respectively the matrices of the explanatory variables and the dependent variable with the observations from week *T*-*m+1*to *T*. Since the true parameter for the price changes from to within the estimation period (i.e., at week ), is not an unbiased estimate of but a weighted average of the true parameters before and after the structural break (e.g., and ). In this example, we assume that there is no structural break after week *T[[2]](#footnote-2)*, and the true demand after week *T* will remain as . Therefore, the forecast error at week *T*+1 becomes:

As a result, the conditional mean of the forecast error will not be zero. The forecast bias is:

=

This can be further illustrated using a simple simulated model. We may 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 determined by the product price but with a structural break at week 31. After the structural break, the product sales become less responsive to the product price reduction. e.g.:

, , when

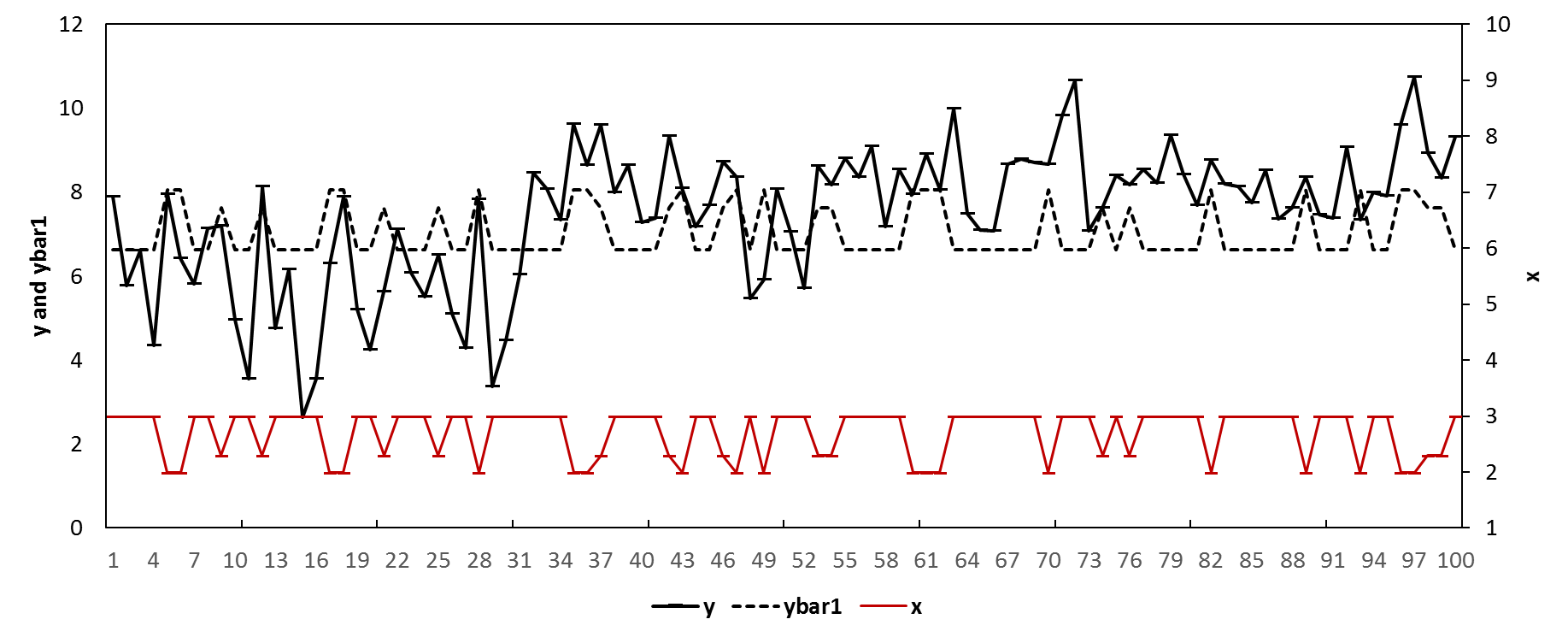
, , when

In this example, the sales become less responsive to temporary price reductions. This may be caused by new product introduction, the intensify of competitive promotional activities by other products, or the change of economic conditions and consumer taste which are unobservable. The sales and price data are illustrated in Figure 1 as the solid black line and the red line respectively.

Suppose we have the data from week 1 to week 70 and we want to forecast the product sales after week 71. This can be illustrated in Figure 1 where the blue area represents the estimation period before the structural break (i.e., [1,30]), the yellow area represents the estimation time period after the structural break (i.e., [31, 70]), and the red area represents the forecast period (i.e., week 71 and after). We may estimate the model with the function form using the data from week 1 to week 70 while ignoring 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 before week 31 and under-predict the product sales for the period from week 31 to week 70, and will produce downwards-biased forecasts for the period after week 70. The predictions/forecasts are represented by the black dashed line in Figure 1. Table 1 shows some error measures which are calculated based on the biased forecasts by this model.

Ideally, if we know that there is a structural break at week 31, we may estimate the model exclusively using the data from week 31 to week 70 and we can generate unbiased forecasts. This is represented by the black dashed line in Figure 2. However, in a retailing context we do not know whether or not there is a structural break and we do not know the location (e.g., week 31 in this example) of the break as the influencing factors are unobservable. Also if the structural break occurs close to the forecast origin and there will not be enough post-break observations to estimate the model.

Figure 1.



Figurer 2

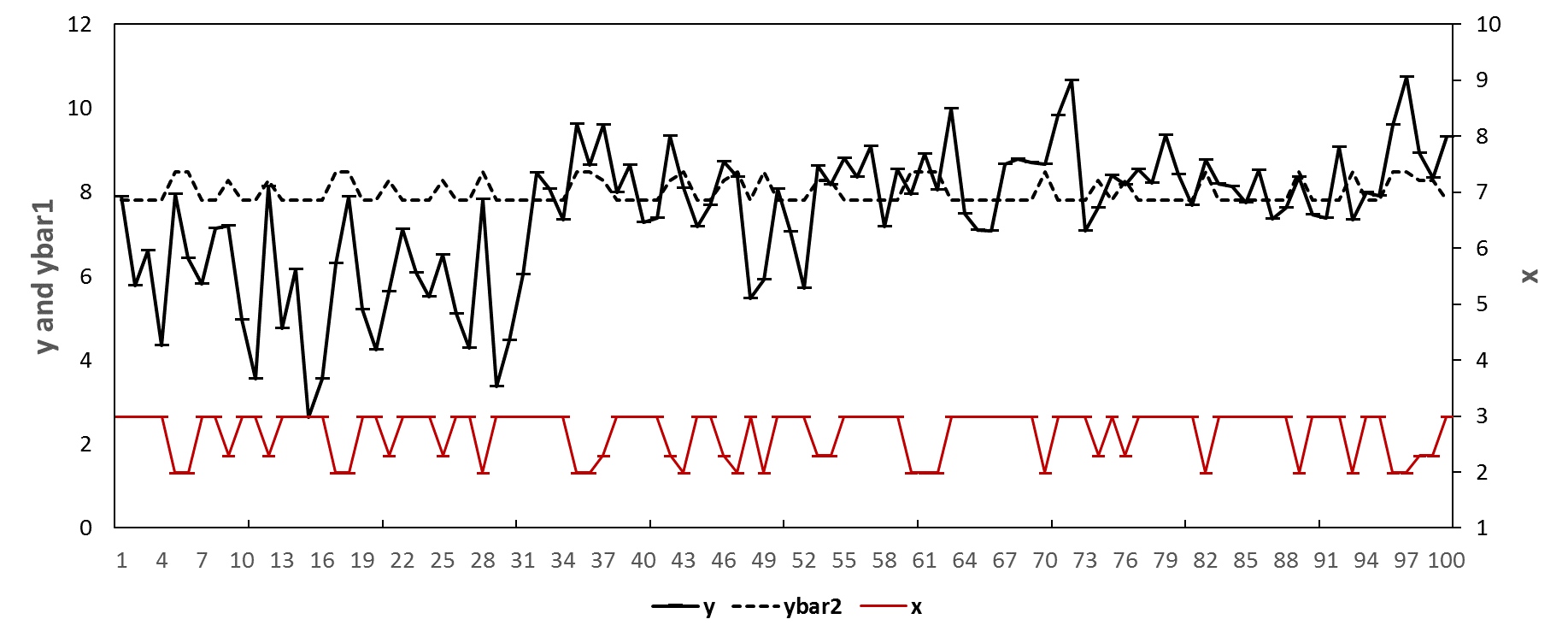


Figure 3.

Table 1. The forecasting performance by the model under differernt scenarios

|  |  |  |  |
| --- | --- | --- | --- |
| 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 full estimation window with intercept correction | 0.824 | 9.77% | 9.54% |
| Figure 4: Model with estimation window combining | 1.034 | 12.17% | 12.58% |

**Section 3: the methods**

**3.1 Intercept correction**

**What is IC**

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_11)). (Clements, 1998; Clements, 1999; Clements, 1998). Specifically, we first try to detect the existence of the structural break. If the model is subject to structural break, we assume that the model generates biased forecasts and we estimate the magnitude of the forecast bias. We then 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, though at the cost of inflated forecasting error variance (*Clements and Hendry 1994 for an example of a simple mode*l).

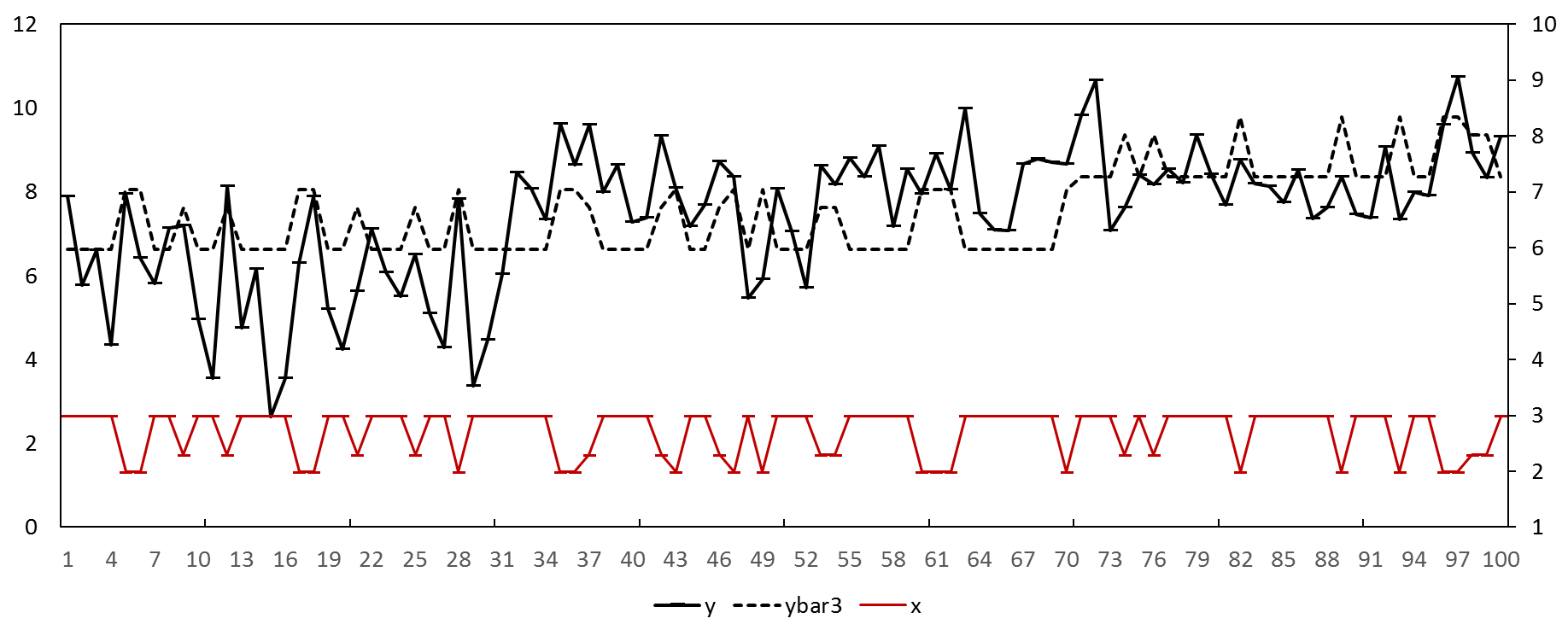
Some analytics about the IC method.

We demonstrate the IC method using the example which was shown the section 3, the model is , with no prior knowledge related to the structural break. We conduct a sequential Chow test based on most of the observations in the estimation period[[4]](#footnote-4). 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 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[[5]](#footnote-5). The results do not suggest the location but only indicate the existence of the structural break. In the literature, different statistic tests have been proposed to detect the locations of the structural breaks (e.g., Chow 1960, Andrews 1993, Andrews and Ploberger 1994, Bai and Perron 1998). However, these tests need to assume priori knowledge of the number of multiple structural breaks.

Figure 3

Based on the results in Figure 3, we confirm 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* is arbitrarily chosen). In this example, we calculate the predictive error for the last four observations in the estimation period. i.e., .

Figure 3.



We then mitigate the forecast bias by adding bias estimate back to the forecasts especially for models with lagged product sales as explanatory variables. Clements and Hendry ([1999](#_ENREF_12)) demonstrated the analytical characteristics of various correction strategies. For example, we could 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. An alternative strategy is to only adjust the one-step-ahead forecast. We could also make adjustments directly to the *h*-step-ahead forecast using the full amount of the forecast bias. In this example, we add the full estimated bias back to the forecasts for each of the forecasts directly. e.g., . The ‘intercept corrected’ forecasts, which are illustrated in Figure 4, are more accurate compared to the forecasts by the original model estimated with the full window regarding various error measures (e.g., MAE= 0.824, MAPE= 9.77%, and SMAPE= 9.54%). The forecasts

The intercept correction method heavily relies on the detection and estimation of the forecasts bias. Also, it 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_12)). Therefore, the contribution by implementing the IC method to conventional models becomes an empirical question.

3.2 Estimation window combining

The intercept correction method may potentially improve the forecasting performance of conventional models by offsetting the forecast bias. An alternative method to improve the forecasting performance of the conventional model is to take a trade-off between the forecast bias and the forecasting error variance by combining the forecast generated by the same model but with different estimation window. 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. Under this condition, the model will not be subject to structural break and the forecasts it generates will be unbiased. However, in reality, the location of the structural break is unobservable. Some statistical tests have been proposed but strong assumptions (e.g., with one single structural break or with known number of structural breaks). Therefore, we could estimate the model with different estimation windows. For example, we may estimate the model with the most recent observations close to the forecast origin and keep the size of the estimation window as small as possible ((we still assume that there are enough observations to estimate the model). The model will hence be less likely to be subject to structural break. For the same simulation example in section 3, we may estimate the model using the data from week 50 to week 70 even when we do not observe the structural break at week 31. The forecasts generated under such circumstance will be unbiased (or less biased, for example, if the estimation window is from week 20 to week 70). We can then combine the forecasts generated by this model and the original model with the full estimation window.

However, the estimation window combining method would not necessarily generate more accurate forecasts as it is associated with a 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 have actually discarded part of the original 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 where the forecast error term is:

[Pesaran and Timmermann (2007)](#_ENREF_49) analyzed the forecasting performance of the model using Mean Square Forecasting Error (MSFE). The MSFE at week *T*+1 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. The change of the MSFE’s when we include one more observation in the model is:

where is the MSFE associated with the model with m+1 observations (the estimation window from week *T*-m to week *T*). It can be proved that the term () is always larger than or equal to zero (i.e., with one more observation before the structural break, the forecast will be more biased), but the sign for the term depends on the percentage of the change of 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 . As a result, 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 such 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.

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

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

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

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

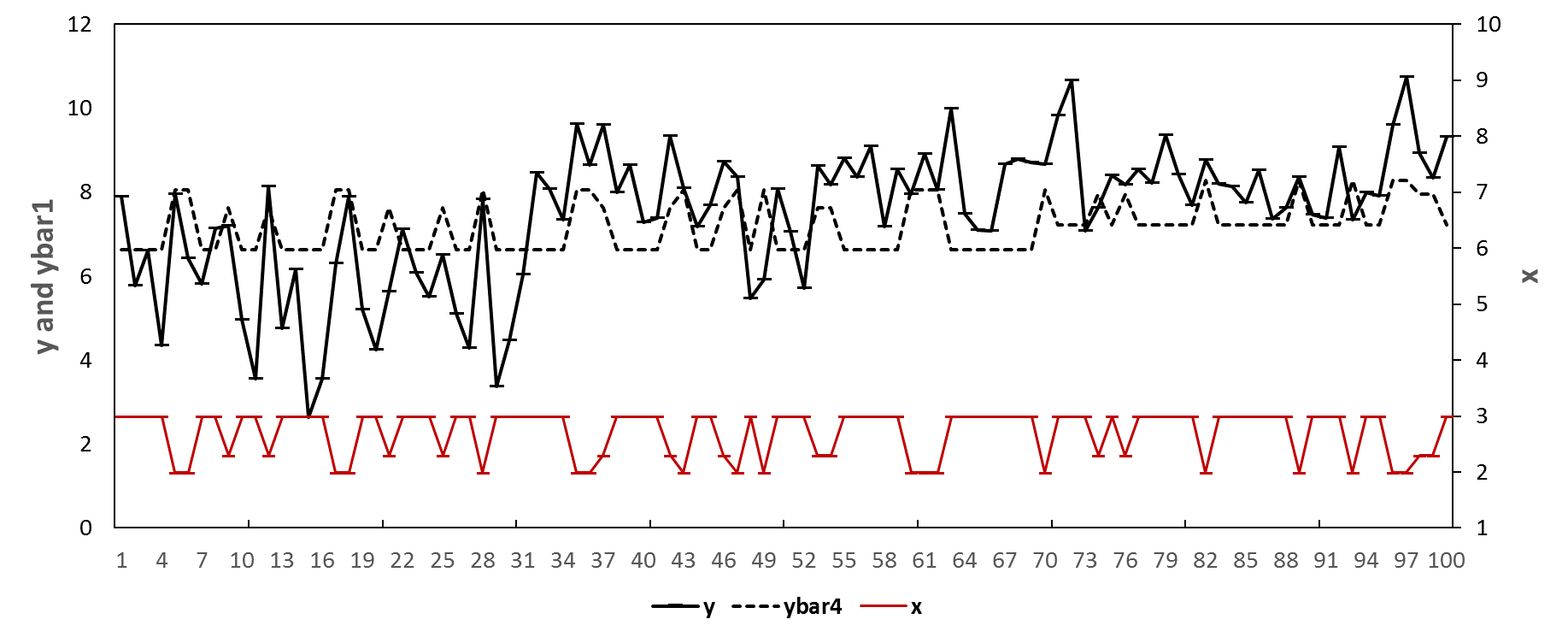
In the combination, can be arbitrarily chosen as long as we can ensure there are enough observations to estimate the model and there are enough variations in all the explanatory variables. Pesaran, Schuermann et al. ([2009](#_ENREF_50)) found that this approach improved the forecasting performance for the random walk with a drift model and the VAR model which are both subject to multiple structural breaks. In this study, we apply the estimation window combining approach with equal weights for the results obtained from each estimation window because it usually generates better performance compared to alternative combining schemes and easy to implement (Stock and Watson, 2001).

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 exact date (e.g., we do not know the parameters change 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 described in section s). Alternatively, we may estimate the model with only the latest data (e.g., the data period from week 41 to week 70)[[6]](#footnote-6). 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 a condition, the forecasts generated by some models (e.g., estimated with the full data, as the one in Figure 1) will be subject to the full bias, 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 subject to less bias (and no bias at 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.

We may mitigate the issue resorting to the forecasting combining studies. We estimate the model with different lengths of estimation window. For example, we estimate 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 will be less biased. This is illustrated in Figure 5. The forecasting accuracy for the EWC method in this simulation example is: 1.034 for MAE, 12.17% for MAPE, and 12.58% for SMAPE.

Figure 4.



The example is clear in the simulation, but in reality many other factors are also involved, for example, variations, how the effectiveness change, missing variables etc. so, do IC and EWC work is an empirical question.

The example is based on a simple static model with one exogenous variables (e.g., the product price). Pesaran and Timmermann (2005) used simulation to demonstrate the rationale.

**Section 4: The data**

In this study, we evaluate forecasting performance of the models using the retail dataset made available by IRI. A descriptive article can be found in [Forni and Reichlin (1996)](#_ENREF_20)[[7]](#footnote-7). The IRI dataset contains weekly data at the SKU level with variables including unit sales, price, features and displays etc. We randomly select 128 SKUs in 15 product categories in one large store[[8]](#footnote-8). Table 1 demonstrates basic statistics for each product category. Some product categories (e.g., Carbonated beverages) have much higher promotion intensity compared to others (e.g., Margarine/Butter and Mayonnaise). Figure 1 illustrates the data for a typical product in the Juice category.

Table 1. 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 1. Unit sales, price (in USD), and promotional events (feature and display) for an SKU in the Beer category.



**Section 5: The models**

In practice, many retailers have been using the base-times-lift method to generate forecasts for their product sales at the SKU level (Cooper et al. 1999). Therefore, we include this method as one of the benchmark models. The method generates the baseline forecast using simple univariate models (e.g., the simple exponential smoothing model) and then makes adjustments for any incoming promotional event:

Where is the baseline forecast for week generated by the simple exponential smoothing model based on the data when there is no promotion. is the actual sales for the previous week when there is no promotion. 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 practice, manufacturers may not be granted the access to retailers’ sales information. Under such circumstance, they need to take advantages of the data they have to achieve a forecasting accuracy as high as possible. Huang et al. (2014) evaluated the forecasting performance of the sophisticated Autoregressive Distributed Lag (ADL) model with price and promotional information of the focal product (i.e., referred as the ADL-own model). Their initial ADL model contains sophisticated dynamics but only engage relatively limited information:

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 *[[9]](#footnote-9)*

are the parameters  
 is the error term and we assume

is the order of the lags[[10]](#footnote-10).

The model is then tested downwards to achieve a valid parsimonious restriction following a general-to-specific modelling strategy. The details can be found in Huang et al. (2014). The model assumes that the effect of the price and promotional information of the focal product to be constant over time while overlooking the impact of the other influencing factors including the impact of competitive price and competitive promotional information within the same product category. In this study, we evaluate the forecasting performance of the ADL-own model when it is combined with the intercept correction method and the estimation window combining method. We refer the combined models as the ADL-own-IC model and the ADL-own-EWC model respectively. These two new models may potentially contribute to improving the forecasting accuracy for manufacturers who do not get access to competitive information from the retailer.

Huang et al. (2014) proposed to incorporate the competitive price and promotional information into the general-to-specific ADL model which significantly improves forecasting accuracy compared to the ADL-own model. There are hundreds of SKU items in the retailing context at the SKU level. They implemented two different methods to keep the model of appropriate size: 1) they select the most relevant competitive price and competitive promotional information using variable selection methods including the Least Absolute Shrinkage and Selection Operator (LASSO) and the stepwise selection; 2) they condense the competitive price and competitive promotional information into a small set of diffusion factors using principle component analysis ([Stock and Watson, 2002)](#_ENREF_52). The initial models corresponding to the ADL model with variable selection methods (referred as the ADL model) 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 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 *[[11]](#footnote-11)*

are the parameters  
 is the error term and we assume

is the order of the lags[[12]](#footnote-12).

The ADL with diffusion factors (referred as the ADL-DI model) is:

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

The ADL model and the ADL-DI model are both subject to downwards testing and simplification following a generate-to-specific strategy. In this study, we evaluate the forecasting performance of these two models when it is combined with the intercept correction method and the estimation window combining method to mitigate the potential structural breaks and/or forecast bias. We refer the combined models based on the ADL model as the ADL-IC model and the ADL-EWC model, and we refer the combined models based on the ADL-DI model as the ADL-DI-IC model and the ADL-DI-EWC model. We expect these models to further improve the forecasting accuracy which has been achieved by retailers.

In this study, the IC method and the EWC method are both implemented in a selective manner based on the existence of structural breaks which are detected via a sequential Chow test. For example, we conduct the Chow ([1960](#_ENREF_10)) test sequentially to investigate whether the model is subject to structural break based on most of the observations in the estimation sample (i.e. 90% central observations). If the null hypothesis of no structural break is rejected for any one of the observations, we consider the model being subject to structural break and forecast bias, and we implement the IC and the EWC method. A very small *p*-value (i.e. 0.005) is used for the sequential Chow test to mitigate the multiple testing problem. For the IC method, we estimate the forecast bias as the (equally weighted) average value of four predictive errors before the forecast origin, and we make direct adjustments to the *h*-step-ahead forecast using the full amount of the forecast bias. For the EWC method, we equally combine the forecasts generated by 40 estimation windows, and we will describe the details in the next section.

Table 1 shows the candidate models:

|  |  |
| --- | --- |
| Models | Description |
| Base-times-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 factors constructed by principle component analysis |
| ADL-DI-IC | ADL-DI model, with IC |
| ADL-DI-EWC | ADL-DI model, with EWC |

For manufacturers: if information is not shared

**Section 6: The experimental design**

In this study, we evaluate the forecasting performance of the candidate models with both fixed origin and rolling origins. For the setting with a 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 typical ordering and planning periods. For the setting with rolling origins, we start with the estimation window from week 1 to week 120, and then re-estimate the model with updated data week by week (e.g., from week 2 to week 121), and so forth. In this study, we conduct 20 sets of rolling experiments and thus we will have 20 sets of one to weeks ahead forecast in total. when we generate the forecast, we use the actual values of the exogenous variables (e.g., price, promotion, or calendar events etc.) the forecasts of the lagged dependent variables when the forecast horizon is beyond one week, For setting with rolling origins, we still specify the ADL variants with the data from week 1 to week 150 assuming a foreknowledge of the data ([Fildes, Wei et al. 2011](#_ENREF_18)). An alternative way is to re-specify the ADL models for each rolling estimation window (Ma et al, 2016). With rolling origins, the results are more robust to randomness and systematic business cycle effects ([Fildes 1992](#_ENREF_17), [Stock and Watson 2002](#_ENREF_53), [Stock and Watson 2002](#_ENREF_52)).

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) proposed by ([Hyndman and Koehler 2006](#_ENREF_27), and the Relative Average Mean Absolute Error (RelAvgMAE) proposed by Davydenko and Fildes (2013). These error measures approximate the utility of the retailer from different aspects.

In this study, the MAPE, the symmetric MAPE, and the MASE for data series *s* with forecast horizon for the rolling event are shown as follows:

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

We compare the overall forecasting performance of the candidate models based on the mean value of all the four error measures across rolling events (for the setting with a fixed origin, *K*= 1) and data series considering different forecasting horizons :

where , , , and are the error measures calculated across data series and rolling events based on forecast horizon (i.e. , , and =1, 4 and 12).

**Section 7: Results**

Section 7.1 fixed origin

Table 1a shows the results for the setting with a fixed forecasting origin for all the 128 SKUs with different forecasting horizons. Under this setting, the Base-times-lift model generally produces the most inaccurate forecasts except when forecast horizon is 1 week ahead. 2) the ADL-own model is outperformed by the ADL model for all scenarios. The performance between the ADL-own model and the ADL-DI model is mixed depending on the forecasting horizon. 3) the ADL-own-IC model and the ADL-own-EWC model both outperform the ADL-own model for all the scenarios. 3) the ADL-EWC model and the ADL-IC model outperform the ADL model for most of the scenarios. 4) the ADL-DI-IC model outperform the ADL-DI model for all the scenarios. The ADL-DI-EWC model have mixed comparative results with the ADL-DI model.

Table 1a. The forecasting performance of the candidate models regarding various forecasting horizons and error measures based on a fixed forecasting origin

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fixed origin, Forecast horizon= 1 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-times-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-times-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-times-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 |

Section 7.2 rolling origins

Table 1a shows the results for the setting with rolling forecast origins for all the SKUs with different forecasting horizons. Under this setting, the Base-times-lift model is outperformed by all the other candidate models for all the error measures and forecast horizons. 2) the ADL-own model is outperformed by the ADL model and the ADL-DI model for all scenarios. 3) the ADL-own-IC model and the ADL-own-EWC model both outperform the ADL-own model for all most of the scenarios. 3) the ADL-IC model outperforms the ADL model for all the scenarios. The ADL-EWC model has mixed comparative results wtih the ADL model. 4) the ADL-DI-IC model outperforms the ADL-DI model for most of the scenarios. The ADL-DI-EWC model has mixed comparative results with the ADL-DI model.

Table 1b. The forecasting performance of the candidate models regarding various forecasting horizons and error measures based on rolling forecasting origins

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Rolling origin, Forecast horizon= 1 | | | | | | | | |
| Models | MAPE | Rank | SMAPE | Rank | MASE | Rank | AvgRelMAE | Rank |
| Base-times-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-times-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-times-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 |

Table 3a shows the forecasting performance of the various models for the time period when the focal product is not being promoted (i.e., there is no feature and display event), based on the setting of rolling origins. Table 3b shows the forecasting performance of the various models for the time period when the focal product is being promoted. The results are consistent with the results for the all the forecast period as described in section 7.2: the base-times-lift has been outperformed by all the other models (Ma et al. 2016); the IC method improves the forecasting performance of the ADL-own method for all the settings for the non-promoted period. The forecasting performance of the ADL-own and the ADL-own-EWC method is mixed for the promoted period.

Table 3a. The forecasting performance of the candidate models regarding various forecasting horizons and error measures based on a fixed forecasting origin, for the non-promoted period.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Rolling origins, the non-promoted period, Forecast horizon= 1 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-times-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-times-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-times-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 3b. The forecasting performance of the candidate models regarding various forecasting horizons and error measures based on rolling forecasting origins, for the promoted period.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Forecast horizon= 1 | | | | | | |
|  | MAPE | Rank | SMAPE | Rank | MASE | Rank |
| Base-times-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-times-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-times-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 |

Section 7.3 forecasting performance for each product category: ADL-own, ADL-own-IC, and ADL-own-EWC

Table 7a investigates how the IC and the EWC method could improve the forecasting performance of the ADL-own model for each of the product categories based on the rolling evaluation. Table 7a compares the forecasting performance of the ADL-own with the ADL-own-IC model and the ADL-own-EWC model for each category and for different forecast horizons with respect to various error measures. The results for the product categories where the ADL-own-IC model and the ADL-own-EWC model respectively outperform the ADL-own model are highlighted in yellow. The results suggest that the ADL-own-EWC model moderately outperform the ADL-own model for MAPE and SMAPE, and has comparative forecasting performance with the ADL-own model for MASE. The improvement by using the EWC method based on the ADL-own model tends to be more significant for longer forecast horizons. The ADL-own-IC model outperforms the ADL-own model for almost all the product categories for all the forecast horizons.

Table 7a xxx

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 7b investigates how the IC and the EWC method could improve the forecasting performance of the ADL model for each of the product categories based on the rolling evaluation. The results for the product categories where the ADL-IC model and the ADL-EWC model respectively outperform the ADL model are also highlighted in yellow. The results suggest that the ADL-EWC has comparative forecasting performance with the ADL model. The ADL-IC model outperforms the ADL model for almost all the product categories for all the forecast horizons.

Table 7b. xxxx

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 7c investigates how the IC and the EWC method could improve the forecasting performance of the ADL-DI model for each of the product categories based on the rolling evaluation. The results for the product categories where the ADL-IC model and the ADL-EWC model respectively outperform the ADL model are all highlighted in yellow. The results suggest that the ADL-EWC model and the ADL-IC model generally outperform the ADL-DI model for most of scenarios.

Table 7c. xxxxx

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |

**Section 8: Conclusion and Future research**

What the is research problem

Grocery retailers needs accurate sales forecasts to effectively manage their inventory.

Why is the problem critically important?

In practice, retailers are facing intense competitions. The profit margin is critical important for retailers. It is difficult to improve the total revenue under such a tight economic condition, so the margin becomes even more crucial. The forecasting accuracy will subsequently impact the performance of the whole supply chain.

How it has been done previously

Some early studies suggested univariate models with adjustments, which help retailer managers directly improve their forecasting procedure. In industry, retailers tend to use a base-lift approach. Later studies proposed more effective forecasting methods. Gur Ali et al. (2009) proposed sophisticated regression tree methods with the price and promotional information of the focal product. Huang et al. (2014) proposed General-to-specific ADL model with competitive price and promotional information within the product category. Ma et al. (2016) designed a three stage ADL model (simplified using a LASSO algorithm) integrating price and promotional information not only within the same product category but also from other product categories. This stream of research aims to improve the forecasting accuracy for retailer product sales at the SKU level by employing more advanced model structures and, more importantly, by incorporating more valuable information which are identified, selected, and refined with advanced techniques (e.g., LASSO and principle component analysis).

How do we contribute with the research in this study

In this study, we contribute to the research problem from a different aspect. Many influencing factors with impact on the retailer product sales are unobservable. For example, the change of the economic conditions [ref, example, UK, Brexit], the change of the consumer taste, new competition entry [example, The German retailer Aldi had an expansion with 400 stores in the United States in the year of 2015 (Loeb, 2015)]. These factors may lead to the change of the effectiveness of the price and promotional activities which have been incorporated in the model. As a result, the model which still assume the corresponding effectiveness of these variables will be subject to structural break and generate biased and less accurate forecasts.

In this study, we employ two different methods including the intercept correction and the estimation window combining to generate more accurate forecasts by dealing with the bias contained in the forecasts of the general-to-specific ADL model. We implement the methods discriminately based on whether the model is subject to structural break using a sequential Chow test. The IC method estimates the magnitude of the forecast bias with the mean values of the predictive errors close to the forecast origin and then add the bias back to the out-of-sample forecasts. The EWC method combines the sets of forecasts by the model with different lengths of estimation windows with equal weights. These two methods have the advantage of mitigating the bias contained in the forecasts produced by the original general-to-specific ADL model but are associated with their own costs. The IC method needs to estimate the magnitude of the forecast bias, which by itself brings in uncertainties. The method adds the estimated bias to the out-of-sample forecasts, which also inflates the forecasting error variance. The EWC method assumes that model with the smallest estimation window is not subject to the structural break. Therefore, whether we can generate more accurate forecasts by taking into account the impact by the unobservable factors is an empirical question.

The results and the implications

In this study, we evaluate the performance of various candidate models in forecasting retailer product sales at the SKU level. Our results have the following indications: 1) the general-to-specific ADL model with the IC method generates the most accurate forecasts over all. Specifically, the ADL-IC model outperforms the original general-to-specific ADL model for all the forecast horizons and error measures. 2) the ADL-DI method also generates more accurate forecasts with the IC method; 3) 2) the ADL-own method also generates more accurate forecasts with the IC method. Manufacturers can benefit from using the ADL-IC method to generate more accurate forecasts when they do not get access to their competitors promotional information from the retailer. 4) the ADL-own model with the EWC method generally has better forecasting performance compared to the ADL-own model, but the ADL-EWC model and the ADL-DI-EWC model have mixed comparative forecasting performance with the ADL model and the ADL-DI model respectively.

Table xx shows the error measure reduction by the IC method and the EWC method.

Our results show that we can improve the forecasting accuracy of the econometric models by using these methods regardless of whether competitive promotional information have been incorporated.

Table xx . the percentage reductions for different forecast horizons and 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% |

Overall, we recommend the ADL-IC model to forecasting retailer product sales at the SKU level as the model generally produces the most accurate forecasts across various scenarios. We also recommend the ADL-own-IC model for manufacturers when competitive promotional information is not available.

Alternatives and future research

The reason for the structural break: there are so many factors which have impact on the product sales and we cannot include all of them. Unless these factors are orthogonal to all the explanatory variables which are already included in the model (e.g., price and promotions), the change of the effect of these factors will lead the model to be subject to structural break and forecast bias.

We tried time-varying parameter models but the model performed poorly. It is difficult to model appropriately the changing process of the promotions. Also the model with varying parameters will have more sophisticated model structure, which may lead to disadvantage for the forecasting performance of the model.

There are alternative method which also mitigate the problem of structural break and forecast bias. [Castle, Doornik et al. (2008)](#_ENREF_9) and [Hendry and Krolzig (2001)](#_ENREF_23) proposed the saturation approach where the regression model was initially incorporated dummy variables for each observation and then recursively reduced by an algorithm called *Autometrics* based on the General-to-specific modelling strategy. The ultimate model will not be subject to structural break and thus would be expected to product unbiased forecast. However the method comes with the cost of losing information (e.g. the observations offset by the retained dummy variables) and its performance becomes an empirical question, and we leave this to our next research step.

IC tries to detect the existence of structural break. There are more advanced structural break tests alternative to the Chow test. There are alternative schemes to estimate the forecast bias. For EWC, there are alternative schemes to combine the forecasts by various estimation windows.

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

We thank the IRI to make the data available.

Reference:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Corsten, D. and T. Gruen (2003). "Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks." International Journal of Retail & Distribution Management **31**(12).

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

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

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

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

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

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

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

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

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

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

Huang, T., et al. (2014). "The value of competitive information in forecasting FMCG retail product sales and the variable selection problem." European Journal of Operational Research **237**(2): 738-748.

Hyndman, R. J. and A. B. Koehler (2006). "Another look at measures of forecast accuracy." International Journal of Forecasting **22**: 679-688.

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

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

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

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).

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.

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

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

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

Monroe, K. B. and J. P. Guiltinan (1975). "A Path-Analytic Exploration of Retail Patronage Influences." The Journal of Consumer Research **2**(1): 19-28.

Morrison, D. G. (1966). "Interpurchase Time and Brand Loyalty." Journal of Marketing Research **3**.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1. 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 is only valid when very restrictive conditions are met. Thereafter in this study, we assume structural breaks lead to forecast bias for the models. [↑](#footnote-ref-1)
2. Clements and Hendry (1999) showed analytically the impact of out-of-sample structural breaks on a VAR model’s forecasting performance. [↑](#footnote-ref-2)
3. This setting is very common in the retailer context. [↑](#footnote-ref-3)
4. 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-4)
5. 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-5)
6. In the simulation, we assume there is only one explanatory variable. Thus 30 observations are more than enough. [↑](#footnote-ref-6)
7. 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-7)
8. We select the SKU’s with positive movements for at least 90% of time. [↑](#footnote-ref-8)
9. We include the 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-9)
10. In the preliminary analysis, *L* is initially set as two. If the general model does not pass the misspecification tests, more lags of the price, promotion, and sales variables are added to the general model. In our modelling, for most UPCs, the ADL models do not contain more than two lags of these variables. [↑](#footnote-ref-10)
11. The calendar events include *Halloween*, *Thanksgiving*, *Christmas*, *New Year’s Day*, *President’s Day*, *Easter*, *Memorial Day*, *4th of July*, and *Labour Day*. [↑](#footnote-ref-11)
12. In the preliminary analysis, *L* is initially set as two. If the general model does not pass the misspecification tests, more lags of the price, promotion, and sales variables are added to the general model. In our modelling, for most UPCs, the ADL models do not contain more than two lags of these variables. [↑](#footnote-ref-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)