**Forecasting retailer product sales in the presence of structural breaks**

Abstract

Grocery retailers need accurate sales forecasts to manage their inventory. We propose more effective methods to forecast their product sales by taking into account the changing effect of the promotional activities. We propose the Autoregressive Distributed Lag (ADL) models with time-varying parameters which try to describe how the effect of the promotional activities change over time. We also implement techniques including estimation window combining and intercept correction to adjust the forecasts by the conventional ADL models. These techniques do not model how the effect of the promotional activities change over time but may potentially improve the models’ forecasting performance by offsetting the forecast bias when the models are subject to structural breaks. The results indicate that we can generate more accurate forecasts by using these two techniques. We evaluate the performance of the various models in a wide range of products.

Key words:

Marketing analytics, Sales Forecasting, promotion

Track:

MODELING AND MARKETING ANALYTICS

Section 1: Introduction

Retailers in the FMCG (Fast Moving Consumer Goods) industry have been struggling with out-of-stock and over-stock for years. When the product is running out-of-stock, retailers not only lose profits but also may lose the customers forever. Recent studies show that customers whom were once believed to either purchase substitutes or postpone their purchases when their preferred products are out of stock are actually more likely to switch stores and never come back ([Corsten and Gruen 2003](#_ENREF_19)). In practice, retailers try to avoid being out-of-stock by deliberately increasing the inventory level (i.e. to over-stock) which inevitably 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).

This challenge has been exaggerated by the promotional activities which intensify the variation of the product sales. Retailers are now facing more intense competition and spending more on promotional activities ([Kamakura and Kang 2007](#_ENREF_34)). Recent studies have proposed sophisticated method with promotional information to generate accurate product sales for retailers. [Gür Ali, SayIn et al. (2009)](#_ENREF_24) introduced machine learning methods including the support vector regression (SVR) and the regression tree method with the promotional information of the focal product. [Huang, Fildes et al. (2014)](#_ENREF_29) developed a two-stage method based on the Autoregressive Distributed Lag (ADL) model which incorporates the promotional information not only from the focal product but also from other competitive products within the same product category. [Ma, Fildes et al. (2016)](#_ENREF_44) proposed a three-stage method based on LASSO modelling strategy which further integrate the promotional information from products in other substitutive/complementary product categories.

These studies, however, are estimated with constant parameters which implicitly assumed that the effect of the promotional activities does not change over time. In practice, this may not be true due to the impact of many influencing factors such as the change of economic conditions, legislation, consumer tastes, and media habits, competition, and advertising etc. ([Wildt 1976](#_ENREF_63), [Wildt and Winer 1983](#_ENREF_64)) As a result, the ADL model with constant parameters are 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_4)). The model being subject to structural breaks may potentially produce biased and less accurate forecasts. In this research, we take into account the change in the effect of the promotional activities and the consequent issue of structural breaks and forecast bias. Specifically, we propose 1) econometric models with time-varying parameters; 2) conventional econometric models with techniques which can offset the potential forecast bias and thus generate more accurate forecasts. We evaluate the performance of these models in forecasting product sales for retailers across a large number of products.

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.

**Section 2: Literature review**

**Previous studies have recognized the change of the effect of the marketing activities.**

In the marketing research literature, a large number of studies have been devoted into understanding the working mechanism of promotional activities (e.g., [Blattberg, Briesch et al. 1995](#_ENREF_8), [Van Heerde, Gupta et al. 2003](#_ENREF_59)). 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_27), [Bijmolt, Heerde et al. 2005](#_ENREF_7)). There are many studies which have been devoted into exploring the changing effects of marketing activities (e.g. [Little 1966](#_ENREF_42), [Morrison 1966](#_ENREF_49), [Myers and Nicosia 1970](#_ENREF_52), [Myers 1971](#_ENREF_51), [Houston and Weiss 1975](#_ENREF_28), [Monroe and Guiltinan 1975](#_ENREF_48), [Moinpour, McCullough et al. 1976](#_ENREF_47), [Wildt 1976](#_ENREF_63), [Wichern and Jones 1977](#_ENREF_62), [Winer 1979](#_ENREF_65), [Mahajan, Bretschneider et al. 1980](#_ENREF_45)). The effect of the marketing activities may change due to exogenous factors, for example, economic condition, legislation, consumer tastes, media habits, competition, and advertising etc. ([Wildt 1976](#_ENREF_63), [Wildt and Winer 1983](#_ENREF_64)).

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_45)). 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_37)). 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_53), [Van Heerde, Srinivasan et al. 2008](#_ENREF_60)).

Intensive promotions can reduce consumers’ reference price ([Lattin and Bucklin 1989](#_ENREF_38), [Lichtenstein and Bearden 1989](#_ENREF_41), [Kalwani, Yim et al. 1990](#_ENREF_33), [Kalwani and Yim 1992](#_ENREF_32), [Foekens, S.H. Leeflang et al. 1999](#_ENREF_23), [Kopalle, Mela et al. 1999](#_ENREF_36), [Levy, Grewal et al. 2004](#_ENREF_40)), 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_61)). 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_39)). 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_46)). 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_6)). 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.

SCAN\*pro model

Foekens, S.H. Leeflang et al. ([1999](#_ENREF_23)) 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_35)) 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_45)). [Cooley and Prescott (1976)](#_ENREF_16) 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_54)), 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).

Main

We can take into account the change in the effect of the promotional activities by develop econometric models with time-varying parameters. Alternatively we can focus on the consequence when the conventional models are subject to structural breaks. There are a large number of studies in the econometrics/economics literature which has been devoted into explore the cause and consequence of structural breaks (e.g. ([Cooper and Nelson 1975](#_ENREF_17), [Clements and Hendry 1998](#_ENREF_14), [Clements and Hendry 1999](#_ENREF_15), [Santos, Hendry et al. 2008](#_ENREF_57)). These studies analytically proved that if the model is subject to structural break, it may generate forecast bias if the structural breaks cause a difference between the deterministic means of the whole estimation sample and the deterministic means at the forecast origin. Techniques have been developed to offset the forecast bias caused by the structural break.

**The challenge caused by structural break**

When the effect of the marketing activities change over time due to the influence of some other factors, conventional econometric models used in some of the early studies mentioned above will be subject to structural break which was defined as large changes in the model’s parameters ([Allen and Fildes 2001](#_ENREF_1)). The parameters estimates of these models were thus a weighted average of the true parameter coefficients before and after the structural break. This may not necessarily generate misleading results for the purpose of policy decision making though it prevents the model showing the whole underlying story. On the other hand, the impact of the structural break on the model’s forecasting performance may be more severe- the forecasts generated by the model will be biased and very likely to be less accurate. The challenge of the structural break on the model’s forecasting performance has been address by many studies especially in the economics literature (e.g. [Cooper and Nelson 1975](#_ENREF_17), [Muellbauer 1994](#_ENREF_50), [Hendry 1995](#_ENREF_26), [Clements and Hendry 1999](#_ENREF_15), [Pesaran and Timmermann 2007](#_ENREF_55), [Castle, Doornik et al. 2008](#_ENREF_11)). [Pesaran and Timmermann (2007)](#_ENREF_55) analytically demonstrated the impact of a structural break within the estimation sample on the model’s forecasting performance. For example, suppose that we have the data for a period of time from week 1 to week *T,* i.e., and we assume the time of a structural break as (). We have the data generating process as a multiple regression as follows:

in this equation, is an indicator which equals to 1 when and 0 otherwise. and are respectively the vectors of the explanatory variables and the dependent variable at time *t*. and are the parameter coefficients before and after the structural break, and we assume that . is the error term, and . We assume that the variance of the error term shifts from to after the time of . We denote that *m* as the first observation in the estimation sample. Under such a circumstance, the model with constant parameters estimated based on the whole sample (i.e., from data *m* to *T*) will be subject to structural break. The corresponding OLS estimates using data from *m* to *T* are:

where and are respectively the matrices of the explanatory variables and the dependent variable with the observations from observation *m* to *T*. The out-of-sample forecasting error at the time of *T*+1 is:

Since is not an unbiased estimate of but a weighted average of and , the expected value of the error, i.e, would not be zero and therefore the forecasting error is not unbiased. In this case, the true DGP will remain as in the out-of-sample data as we explicitly assume that there is no structural break after the time *T[[1]](#footnote-1)*.

If we know the exact time (i.e. ) for the structural break, we may simply estimate the model with the data after the structural break (i.e., ). The model will not be subject to the structural break and the forecast will be unbiased. However, this strategy has two limitations. First, it comes with a cost of inflated forecasting error variance. Second, is usually unknown in practice. We will explain these in more details in section x.

**Section 3: Methodology**

**Model1: TVP**

Some early studies in the marketing literature allowed the parameters of the marketing activities to change over time. These studies have adopted models with parameters of different types of variations in forecasting product sales. One type of the models is the systematic variation models where the parameters of the marketing activity variables are related to exogenous environmental factors and/or previous marketing activities. The model can be represented as , where is the product sales, is a vector of the marketing activity variables, and is the error term. is a vector of parameter coefficients which is modelled as a function of a constant term (i.e. ), exogenous environmental variables (i.e. ), and an error term (i.e. ). is assumed to unrelated with . If we assume the functional form *f* to be linear, can be written as , is the parameter vector. Mahajan, Bretschneider et al. ([1980](#_ENREF_45)) used similar functional form to model the effects of advertising, and they allowed the effects of the advertising to change over time because the design and the content of the advertising were changed during the sample period. However, early studies showed that the system variation models could exhibit poor forecasting performance when the model makes inappropriate assumptions about the specification of time-varying parameters ([Helmer and Johansson 1977](#_ENREF_25)).

Some other studies have tried to model the parameters of the marketing activity variables such as advertising and price with autoregressive variations and random variations. Jex ([1994](#_ENREF_31)) proposed a discount weighted regression (DWR) method, allowing the parameters to follow an autoregressive form, to estimate the multiplicative market share model proposed by Brodie and De Kluyver ([1987](#_ENREF_9)). Cooley and Prescott ([1976](#_ENREF_16)) proposed an alternative model where the parameters is assumed to evolve overtime as , , where and are the error terms. Models with parameters of autoregressive variations have been applied to capture how the effects of marketing mix variables change over time (e.g. [Little 1966](#_ENREF_42), [Pekelman and Edison 1980](#_ENREF_54), [Liu and Hanssens 1981](#_ENREF_43)).

In this study, we propose the models with time-varying parameters of stochastic variations. We give the models more flexibility to capture how the effect of the promotional activities change over time. The models may potentially be more robust to structural break and thus generate more accurate forecasts by reducing the forecast bias. We allow the parameter coefficients of the explanatory variables to follow a simple autoregressive process (i.e. AR(1)). Thus if we have an original econometric model as , we will then have the parameters as , and , where and are respectively the explanatory variables and the dependent variable. and are the vectors of the parameters at time , . , and are the error terms. In this model, and are assumed to evolve over time, and only their latest estimate (i.e., calculated with more weights on the most recent observations) will be used to generate the out-of-sample forecasts, which may potentially accounts for the change in the effects of the promotional activities.

**Model2: EWC**

If we know the model is subject to structural break, we could simply discard the data before the structural break if we know the exact date of the break.

When the model is subject to structural break, a conventional approach is to estimate the model using the data after the structural break. For example, suppose we have the DGP as described in section 3 that , for the data of . If the time of the structural break is known, the model could simply be estimated based on the data after the break, i.e., [1: *T*] and the model will not be subject to the structural break. The first limitation of this strategy is that we usually do not know the location of the structural break (i.e., ). We could only rely on statistic tests (e.g., [Chow 1960](#_ENREF_12), [Andrews 1993](#_ENREF_2), [Andrews and Ploberger 1994](#_ENREF_3), [Bai and Perron 1998](#_ENREF_5)), but the potential existence of multiple structural breaks makes the test even more challenging. However, even we know the location of the structural break, we may still need to include the pre-break data because we may not have enough observations to estimate the model if the structural break occurs close to the end of the estimation period.

Alternatively, [Pesaran and Timmermann (2007)](#_ENREF_55) proposed a very different approach. They proved analytically that including the data before the structural break would inevitably make the forecasts to be biased, but potentially make the forecasts to be more accurate under certain circumstance. Suppose we have the forecasting error as described in section 3:

Thus the forecasting error measure Mean Square Forecasting Error (MSFE) at the time of *T*+1 conditional on is as follows ([see equation (7) in Pesaran and Timmermann 2007](#_ENREF_55)):

where , and is a diagonal matrix where the first diagonal places are and the remaining diagonal places are . The can be decomposed as ([see equation 8 in Pesaran and Timmermann 2007](#_ENREF_55)):

where

In this equation, is the squared forecast bias, and is the efficiency term ( is the forecasting error variance). Thus we can examine how the changes when one additional observation is added in the estimation sample. The change in the is defined as ([see equation 15 in Pesaran and Timmermann 2007](#_ENREF_55)):

where is the MSFE calculated with an estimation window with one extra observation compared to . Pesaran and Timmermann ([2007](#_ENREF_55)) showed that the term () is always larger than or equal to zero, and the sign of depends on the sign of (i.e. ). For example:

where and

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

Pesaran and Timmermann ([2007](#_ENREF_55)) proposed to estimate the model with a large number of estimation windows with or without pre-break data and then combine the forecasts generated with these estimation windows, which resorts to the philosophy of forecasting combination. 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_56)) 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. There could be various combining schemes including equal weight average, exponential weighted average, and Bayesian combining etc., and even the optimal combining scheme could be developed given the location of the structural break. In this study, we apply the estimation window combining technique with the equal weights because it usually generates better performance compared to alternative combining schemes and easy to implement (Stock and Watson, 2001).

**Model3: IC**

Another technique is the intercept correction (IC) which offsets the forecast bias by specifying non-zero values for the model’s errors in the forecast period. Given that structural breaks are detected, the technique will estimates the magnitude of the forecast bias caused by the structural break and then add the value of the bias back to the out-of-sample forecasts. The intercept correction technique may potentially improve the forecasting accuracy by mitigating the forecast bias. The technique has been widely used in making adjustments for macro-economic forecasts ([Clements and Hendry 1994](#_ENREF_13)).

The forecast bias can be estimated using different schemes. In this study, we estimate the forecast bias as the average value of four predictive errors before the forecast origin.

Another approach to take into account the structural break is the intercept correction (IC). The technique has been widely used in making adjustments for macro-economic forecasts ([Clements and Hendry 1994](#_ENREF_13)). It attempts to offset the forecast bias by specifying non-zero values for the model’s forecasting errors. For example, if any structural break is detected in the estimation period, the intercept correction technique will first estimate the forecast bias caused by the structural break. This can be done with various estimation schemes. One usually scheme is to estimate the forecasting bias as the predictive error at the forecast origin (i.e., , where *T* is the last observation in the estimation window). Alternatively, the forecasting bias could be estimated as the average value of an ad hoc number of predictive errors before the forecast origin. (e.g. , where *i* is arbitrarily chosen).

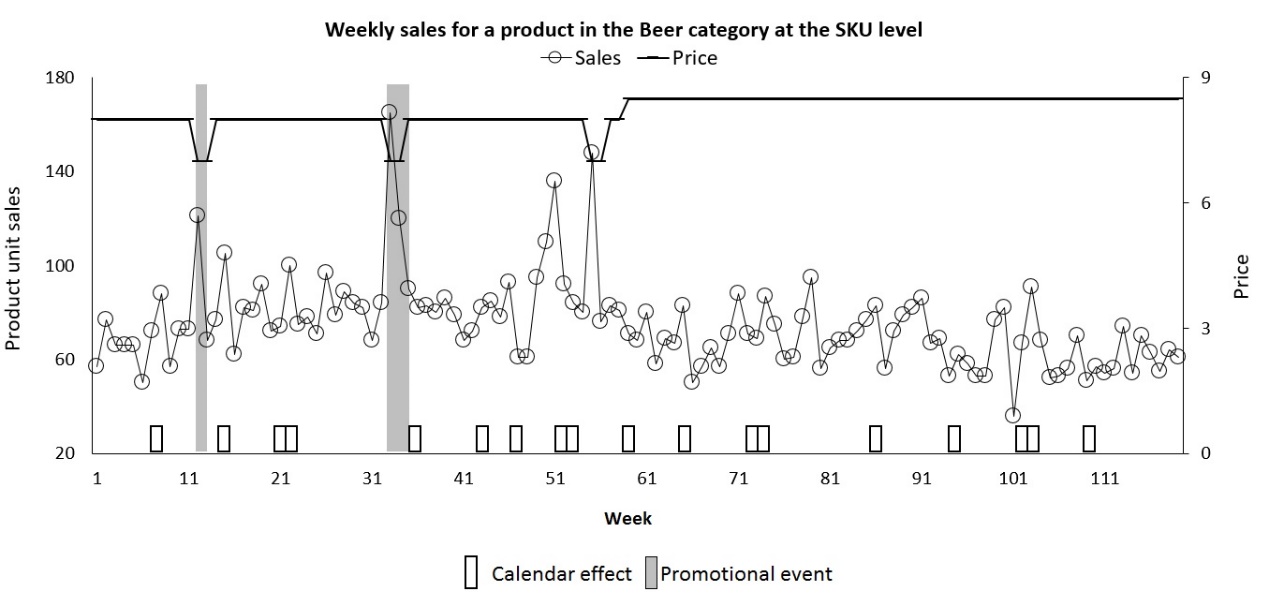
The approach will then add the estimated bias back to the out-of-sample forecasts following various correction strategies. Clements and Hendry ([1999](#_ENREF_15)) demonstrated the analytical characteristics of various correction strategies using an example of VAR(1) model with a time trend, i.e., . Suppose that the model is subject to structural break and the forecast bias is estimated as . Denote , , and , as the corrected *h*-step-ahead forecast by the intercept correction technique following various strategies. The intercept correction approach could first makes adjustments to the one-step-ahead forecast, and then calculate the two-step-ahead forecast based on the value of the one-step-ahead forecast which has already adjusted, and so forth. The adjusted *h*-step-ahead forecast is described as . This equation can be re-written recursively as , where is the original *h*-step-ahead forecast. An alternative strategy is to only adjust the one-step-ahead forecast, and . This equation can be re-written recursively as . Another correction strategy makes adjustments to the *h*-step-ahead forecast using the full amount of the forecast bias. That is, .

The intercept correction technique can potentially reduce the forecasting bias if the forecasting bias is estimated properly. However, whether it could improve the forecasting accuracy is an empirical question because this approach comes with the cost of inflated forecasting error variance. Clements and Hendry ([1999](#_ENREF_15)) derived analytically the forecast bias and the inflated forecasting error variance for the VAR(1) model described the various correction strategies. Their findings show that these correction strategies have their own advantages and limitations in terms of the reduced forecasting bias and inflated forecasting error variance depending on the details of the structural break such as which of parameters have changed and whether there are multiple breaks etc.

In this study, we estimate the forecast bias as the (equally weighted) average value of four predictive errors before the forecast origin, and we implement the intercept correction approach using the last correction strategy. Also the intercept correction approach is implemented in a discriminated manner considering the existence of the structural break. For example, we first conduct the Chow ([1960](#_ENREF_12)) test sequentially to investigate whether the model is subject to structural break based on most of the observations in the estimation sample (i.e. over 90% central observations). If the test is rejected for any of the observations, the model is identified as being subject to structural break. A very small *p*-value (i.e. 0.005) is used to mitigate the multiple testing problem in detecting the existence of the structural break. The intercept correction technique will be implemented if and only if the model is identified as being subject to structural break.

Section 4: Data and Experimental design

We use the data from the IRI company for which a descriptive article can be found in [Bronnenberg, Kruger et al. (2008)](#_ENREF_10)[[2]](#footnote-2). The IRI dataset contains weekly data at the SKU level including information on sales, price, features and displays. We select 300 SKUs in 30 product categories in a large store. The SKUs we include in our experiment all have positive movement for at least 90% of time.



**Following the first two stages modelling developed by Huang et al. (2014)**

In this study, we propose three-stage models. For example, we select/summarize the competitive promotional information, and then develop the general-to-specific ADL models following Huang, Fildes et al. (2014). At stage three, we develop time-varying parameter models based on the ADL models, and we also implement techniques including the estimation window combining (EWC) and intercept correction techniques (IC) to the ADL models. We briefly introduce the first two stage as follows:

At the first stage we conduct the variable selection method and the PCA analysis as described in Huang, Fildes et al. (2014). Huang, Fildes et al. (2014) proposed to 1) to select the most important promotional variables using variable selection algorithms; 2) to summarize the general information across all the promotional variables using the PCA (principle component analysis at an acceptable cost of information loss.

At the second stage we develop general-to-specific ADL model with the competitive promotional information. The ADL models are developed based the explanatory variables retained by the variable selection methods at stage one. The variables retained by the stepwise regression and the LASSO algorithm are used in combine to avoid missing important explanatory variables (though at the cost of loss in efficiency due to over-specification). Alternatively, the ADL models are developed with the constructed Diffusion Factors at stage one. The details of the developing the ADL models can be found in Huang, Fildes et al. (2014).

**The ADL model**

week before the event if . takes the values from 1 to 9 representing all the calendar events *[[3]](#footnote-3)*

are the parameters  
 is the error term and we assume

is the order of the lags[[4]](#footnote-4)

The general ADL model will ideally pass all the misspecification tests (e.g. the *F*-test, the Breusch-Godfrey test for autocorrelation, and tests for heteroskedasticity and normality). The model may be estimated by OLS with the usual interpretations of the statistics whether or not the data series are stationary, because sufficient lags were included to remove any autocorrelation (although with some potential loss of efficiency) ([Song and Witt 2003](#_ENREF_58)). A well-specified ADL model can then be simplified following the general-to-specific strategy. For example, we first estimate the general ADL model and remove the explanatory variable with the highest *p*-value for the parameter restriction test. We then estimate the reduced model and re-conduct all the misspecification tests. If the reduced model passes all these tests, we move on to remove the variable with the highest *p*-value in the new estimation, provided that the previous variable has already been removed, and so forth. Otherwise we will add the variable back and repeat the process by removing the variable with the second highest *p*-value. In the modelling process we remove the variables with incorrect signs. We also remove the explanatory variables which are not economically significant (i.e. with very small parameter coefficients) to achieve parsimony. The final simplified ADL model must pass all the misspecification tests which are passed by the general ADL model. The model is estimated by OLS with robust estimators in the presence of heteroscedasticity. The following example shows the general-to-specific ADL model with the diffusion indexes:

where

is the diffusion index of competitive prices at week .

is the diffusion index of competitive promotion at week .

*P* and *Q* are the numbers of initially retained diffusion indexes, and [[5]](#footnote-5)

**Following the experimental design by Huang et al. (2014)**

In this study, we implement the base-times-lift approach as one of the benchmark models as it is usually used by retail practitioners ([Cooper, Baron et al. 1999](#_ENREF_18), [Gür Ali, SayIn et al. 2009](#_ENREF_24)). The model can be represented as:

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

In this study, we evaluate the models’ forecasting performance based on a rolling scheme following previous studies (e.g., [Huang, Fildes et al. 2014](#_ENREF_29), [Ma, Fildes et al. 2016](#_ENREF_44)). The rolling evaluation ensure results are more robust to randomness and systematic business cycle effects ([Fildes 1992](#_ENREF_21)). Specifically, we evaluate the models with 30 rolling forecast origins with multiple forecast horizons. For example, we estimate the models with a moving window of 120 weeks and forecast one to weeks ahead where is 1, 4, and 12 chosen by taking into account typical ordering and planning periods. We then re-estimate the model with updated data by including the data in the latest week and dropping the data in the earliest week. We repeat this process until we use all the data in the remaining estimation sample. We have 30 sets of one to weeks ahead forecast. When the lead times are greater than one, the actual value of the explanatory variables and the forecasted values of the lagged dependent variables are used. In practice, promotional variables are usually known to retailers as they are included as one part of the agreed promotional plan between retailers and suppliers. We develop the ADL models based on the data from week 1 to week 150 to represent the model which would ideally be specified with the foreknowledge of the data ([Fildes, Wei et al. 2011](#_ENREF_22)). An alternative way to evaluate the models is to re-specify the model for each rolling event based on each moving estimation window ([e.g., Ma, Fildes et al. 2016](#_ENREF_44)).

We follow Huang et al. (2004) to evaluate the models’ forecasting performance using five error measures: the MAE, the Mean Absolute Scaled Error (MASE), the MAPE, the symmetric Mean Absolute Percentage Error (sMAPE), and the Average Relative Mean Absolute Error (AvgRelMAE). In this study, the MAE for data series *S* calculated with forecast horizon for the rolling event is:

where is the actual value in the forecast period for data series based on the rolling event, and is the forecast value for data series based on the rolling event[[6]](#footnote-6).

The MASE, proposed by [Hyndman and Koehler (2006)](#_ENREF_30), represents the “weighted” arithmetic mean of the MAE compared to the variations in the estimation sample. The MASE calculated across data series with forecast horizon for the rolling event is as follows:

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

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

However, both percentage error measures including the MAPE and the sMAPE can be distorting when the actual values and the forecast values are relatively small compared to the forecast error, in which case the resulting percentage errors become extremely large ([Hyndman and Koehler 2006](#_ENREF_30)). The sales at the UPC level exhibit high degree of variations due to seasonal effects, changing stages of product life cycle, and particularly promotional activities. Under such a circumstance, it is very likely to have large forecast errors associated with relatively low product sales, which makes the percentage based error measures less advisable in our context ([Davydenko and Fildes 2013](#_ENREF_20)).

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

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

Considering the limitations of the four error measures, [Davydenko and Fildes (2013)](#_ENREF_20) recommended the AvgRelMAE, which is a geometric mean of the ratio of the MAE between the candidate model and the benchmark model. In this study, we take the average of the AvgRelMAE across all the rolling events (i.e. ) and S data series (i.e. ) with respect to forecast horizon :

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

In this study, we follow the experimental design by (Huang, Fildes et al. 2014). That is, 70 rolling events with 120 weeks estimation window for a 12 weeks ahead forecasting horizon. We specify the variants of ADL models with the data from week 1 to week 200, which represents the model that would ideally be constructed based on a foreknowledge of the data (Fildes, Wei et al. 2011). Alternatively, as in Ma, Fildes et al. (2016), the models were re-specified for each rolling event based on each moving estimation window, though they have used a different modelling strategy. We have the following candidate models:

|  |  |
| --- | --- |
| Base-times-lift | Industrial practice, simple-exponential smoothing with adjustments based on the effect of the most recent promotion |
| ADL-own | ADL model, with the promotional variables of the focal product |
| ADL | ADL model, based on the variables retained by stepwise and LASSO |
| ADL-DI | ADL model, based on the diffusion factors |
| ADL-EWC | ADL model with estimation window combining |
| ADL-DI-EWC | ADL-DI model, with estimation window combining |
| ADL-IC | ADL model with intercept correction |
| ADL-DI-IC | ADL-DI model, with intercept correction |
| TVP | ADL model with time varying parameters modelled as an AR(1) process |

In this study, the intercept correction technique is implemented in a selective manner considering the existence of structural breaks. We first conduct the Chow test sequentially to investigate if the model is subject to structural breaks. The test was conducted for each observation in the estimation period. If the test is rejected for any one of the observations, the model is considered as being subject to structural breaks. A very small p-value (i.e. 0.005) is used to mitigate the multiple testing problem in detecting the structural break. The estimation window combining technique and the intercept correction technique will be implemented if and only if the model is identified as being subject to structural breaks.

In this study we evaluate the forecasting performance of the various models using three error measures: the Mean Absolute Scaled Error (MASE), the MAPE, the symmetric Mean Absolute Percentage Error (sMAPE), each of which are calculated across rolling events and data series. We test the statistical significance for the difference between the forecasting results of the various models using the Wilcoxon signed rank (SR) test which can be considered as a non-parametric version of a paired sample t-test but does not assume the errors follow any specific distribution.

Section 5: Results

We have the results by all the candidate models for the seven product categories overall in terms of different forecast measures. We also examine the forecasting performance of the models for the time period considering if the focal product is on promotion. The models will have the same ranks if the difference in their performance measures is not statistically significant. The results in Table 1 indicate: 1) the EWC techniques and the IC techniques generally and substantially improve the performance of the ADL model; 2) the TVP model has poor forecasting performance. This may indicate that the change in the effect of the promotional activities has not modelled well by an AR(1) process; 3) the ADL-DI-IC model has the best forecasting performance across all the products.

Table 1 forecast results based on the 1-12 weeks forecast horizon

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Models | MASE | Rank | SMAPE | Rank | MAPE | Rank |
| all forecast period |  |  |  |  |  |  |
| Base-times-lift | 0.806 | 9 | 32.7% | 9 | 32.0% | 7 |
| ADL-own | 0.666 | 7 | 25.6% | 7 | 32.5% | 7 |
| ADL | 0.643 | 5 | 23.8% | 6 | 28.4% | 6 |
| ADL-DI | 0.632 | 2 | 23.0% | 2 | 27.3% | 2 |
| ADL-EWC | 0.638 | 1 | 23.3% | 2 | 27.5% | 1 |
| ADL-DI-EWC | 0.633 | 1 | 22.8% | 2 | 26.8% | 1 |
| ADL-IC | 0.627 | 1 | 23.0% | 2 | 26.6% | 1 |
| ADL-DI-IC | 0.623 | 1 | 22.5% | 1 | 25.9% | 1 |
| TVP | 0.719 | 8 | 26.5% | 8 | 32.0% | 7 |
| Promoted forecast period | |  |  |  |  |  |
| Base-times-lift | 1.808 | 9 | 55.1% | 9 | 41.7% | 9 |
| ADL-own | 1.202 | 1 | 29.1% | 1 | 34.9% | 7 |
| ADL | 1.194 | 1 | 28.1% | 1 | 31.8% | 1 |
| ADL-DI | 1.157 | 1 | 27.3% | 1 | 31.5% | 3 |
| ADL-EWC | 1.242 | 1 | 28.3% | 1 | 32.1% | 1 |
| ADL-DI-EWC | 1.201 | 1 | 27.6% | 1 | 31.7% | 1 |
| ADL-IC | 1.191 | 1 | 27.8% | 1 | 30.6% | 1 |
| ADL-DI-IC | 1.147 | 1 | 27.1% | 1 | 30.4% | 1 |
| TVP | 1.393 | 8 | 33.5% | 8 | 39.8% | 8 |
| Non-promoted forecast period | | |  |  |  |  |
| Base-times-lift | 0.497 | 6 | 22.3% | 6 | 27.8% | 6 |
| ADL-own | 0.492 | 6 | 24.1% | 6 | 31.1% | 6 |
| ADL | 0.467 | 6 | 21.7% | 6 | 26.2% | 6 |
| ADL-DI | 0.453 | 1 | 20.6% | 1 | 24.3% | 1 |
| ADL-EWC | 0.456 | 1 | 21.0% | 1 | 24.8% | 1 |
| ADL-DI-EWC | 0.447 | 1 | 20.3% | 1 | 23.8% | 1 |
| ADL-IC | 0.449 | 1 | 20.7% | 1 | 24.2% | 1 |
| ADL-DI-IC | 0.445 | 1 | 20.0% | 1 | 22.9% | 1 |
| TVP | 0.495 | 6 | 23.5% | 6 | 27.5% | 6 |

Future research,

Clements and Hendry, out-of-sample structural break and forecast bias

OxMetrics

Reference:

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

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

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

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

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

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

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

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

Brodie, R. J. and C. A. De Kluyver (1987). "A comparison of the short term forecasting accuracy of econometric and naive extrapolation models of market share." International Journal of Forecasting **3**(3-4): 423-437.

Bronnenberg, B. J., et al. (2008). "The IRI Academic Dataset." Marketing Science **27**(4).

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 (1998). "Intercept corrections and structural change." Journal of Applied Econometrics **11**: 475-494.

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).

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

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

Davydenko, A. and R. Fildes (2013). "Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts." International Journal of Forecasting(0).

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.

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

Helmer, R. M. and J. K. Johansson (1977). "An Exposition of the Box-Jenkins Transfer Function Analysis with an Application to the Advertising-Sales Relationship." Journal of marketing research **14**(2): 227-239.

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

Jex, C. F. (1994). "Recursive estimation as an aid to exploratory data analysis: an application to market share models." International Journal of Forecasting **10**(3): 445-453.

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.

Kamakura, W. A. and W. Kang (2007). "Chain-wide and store-level analysis for cross-category management." Journal of Retailing **83**(2): 159-170.

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).

Liu, L.-M. and D. M. Hanssens (1981). "A bayesian approach to time-varying cross-sectional regression models." Journal of econometrics **15**(3): 341-356.

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.

Santos, C., et al. (2008). "Automatic delection of indicators in a fully saturated regression." Computational Statistics **23**: 317-335.

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

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

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

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

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

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

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

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

1. Clements and Hendry (1999) showed analytically the impact of out-of-sample structural breaks on a VAR model’s forecasting performance. [↑](#footnote-ref-1)
2. 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-2)
3. 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-3)
4. 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-4)
5. We choose to retain four diffusion indexes for competitive prices and four diffusion indexes for competitive promotions based on preliminary analysis. For each product category, the percentages of explained variation in the competitive price data series range from 51% to 79%, and the percentages of explained variation in the competitive promotion data series range from 32% to 69%. [↑](#footnote-ref-5)
6. 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-6)