In the second stage, we incorporate the refined competitive information into econometric forecasting models. In this study, we construct the Autoregressive Distributed Lag (ADL) model following a general-to-specific modelling strategy ([Hendry, 1995](#_ENREF_37)). We choose the ADL model for several reasons. First, the ADL model has the advantage of taking into account the carryover effect of the price and promotional variables. Second, the general-to-specific modelling strategy ensures the parsimony and data congruence of the model. Third, the ADL model is transparent with a simple regression style model structure, which benefits the users ([Fader and Hardie, 2005](#_ENREF_21)). It has good interpretability compared to “black box” machine learning approaches which can hardly be understood by brand/category managers. Also, in the forecasting literature, the general-to-specific ADL model is one of the most popular time series forecasting models and has exhibited superior forecasting performance in other areas including manufacturer sales, tourism, and air passenger flows (see [Albertson and Aylen, 2003](#_ENREF_3); [Fildes et al., 2011](#_ENREF_26); [Song and Witt, 2003](#_ENREF_61)). Specifically, we start with a general model assuming that it properly describes the salient features of the data generating process, and then simplify the general model by seeking out valid parsimonious restrictions. The following example shows the general ADL model with the most relevant competitive explanatory variables identified by the stepwise selection and the LASSO selection procedure:

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 *[[1]](#footnote-1)*

are the parameters  
 is the error term and we assume

is the order of the lags[[2]](#footnote-2).

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_61)). 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 for the parameter restriction test. In the modelling process we also remove the variables with incorrect signs and those not economically significant (i.e. with very small parameter coefficients) to achieve parsimony. The final simplified ADL model must pass all the misspecification tests of the general ADL model. The model is estimated by OLS with robust estimators in the presence of heteroscedasticity. Analogously, the following example shows the general-to-specific ADL model with the diffusion indexes:

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

1. 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-1)
2. 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-2)