Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 50-1 (2017) 13003–13008

# Control-oriented modeling of SKU-level demand in retail food market

**M. Mazzoleni** *, **S. Formentin** **, **F. Previdi** *, **S. M. Savaresi** **

* *Department of Management, Information and Production engineering University of Bergamo, via Galvani 2, 24044 Dalmine (BG), Italy (e-mail: mirko.mazzoleni@unibg.it).*
** *Department of Electronics, Information and Bioengineering, Politecnico di Milano, via G. Ponzio, 34/5 - 20133 Milano (MI), Italy*

**Abstract:** In food market, modeling the dynamics of Stock-Keeping Unit (SKU) requests is of fundamental importance, not only to understand the market but also for optimization and control purposes. In fact, standing on model-based predictions of future demand, an efficient planning of the promotional calendar can be devised. Moreover, better inventory management can be achieved, by reducing losses due to expired aliments remained unsold and improving distribution operations. In this work, data-driven control-oriented modeling of such a demand is discussed and a novel switching dynamical strategy is proposed. When applied to experimental data from a real food company, the above strategy is shown to accurately predict future sales under fixed promotion events and outperform the state-of-the-art modeling methods.



*Keywords:* Knowledge discover (data mining), Identification and model reduction, Production planning and control

## 1. INTRODUCTION

The ability to accurately forecast the future demands of goods and services gives a clear competitive advantage over competitors, leading to higher profits by constantly optimizing operation management tasks, such as inventory management, planning and scheduling, as described in Samson and Singh (2008). In the context of food retail market, Stock-Keeping Unit (SKU) request prediction is of paramount importance to reduce losses caused by expired aliments that remained unsold, enhance customer satisfaction and improve distribution operations. One of the main factors contributing to sold quantity variation is the employment of promotional events by the retailers. Usually these are planned collaboratively by them and manufacturers, who jointly agree on the products, types, price reduction and the timing of promotions. Products are typically on promotion for a limited period of time, in the order of days, during which demand is usually substantially higher than during periods without promotions, see Cooper et al. (1999). Here, the authors also identified that the display location of items in retail stores, weather and holiday periods had a positive impact on sales. This dramatical change in sales behaviour (*uplift*) with respect to the quantity sold during non-promotional days (*baseline*) makes standard inventory management and replenishment techniques (e.g. based on reorder point/order quantity policies) not suitable for an efficient management. There is therefore a need for more intense collaboration, such as Collaborative Planning, Forecasting and Replenishment (CPFR) procedures, see Ireland and Crum (2005); Ailawadi et al. (2009).

However, despite the benefits of CFPR, collaboration and information sharing does not prevail, as shown in Holweg et al. (2005). When collaborative information is not available, the supplier must rely on historical data and qualitative knowledge of the marketplace to build the forecast. The effectiveness of data-driven decisions on firm performance has been assessed in Brynjolfsson et al. (2011). To tackle this problem, two basic approaches have been considered: judgemental and quantitative forecasting, see Armstrong (2001). A combined approach, aiming at supporting human decisions on predictions provided by the statistical model, has been taken into consideration in Lee et al. (2007). Some of the most used models in literature consist of the Simple Exponential Smoothing (SES), the "last-like promotion" and the autoregressive distributed lag (ADL) model. The SES model, shown to be efficient in capturing the level component in demand over time, see Huang et al. (2014), does not make use of any promotional information: when the characteristics of the demand series change due to special events such as promotions or holidays, the fitted parameters can no longer describe the series. The so-called "last-like promotion" model, as implemented in Ali et al. (2009) is one of the simplistic models that use exogenous variables. This method first generates a baseline forecast using a model, such as the SES, for non-promoted time periods. A "lift effect" is then added to the baseline forecast during the promoted periods. The autoregressive distributed lag (ADL) model, explained in Huang et al. (2014), considers instead various types of predictors, such as: past values of demand of the considered product, price of the considered product, promotional index of the considered product, price of the competitors products, promotional indices of the competitors products, monthly indicator variable and


10.1016/j.ifacol.2017.08.1951

calendar events. Many studies have focused on evaluating the effect of promotions on sales, using store or market level data, see Blattberg and Levin (1987); Abraham and Lodish (1987). Yang et al. (2015) proposed a hierarchial model at single Universal Product Code (UPC) level, while Banek et al. (2015) investigated the presence of outliers in developing promotional sale forecasting methods. The work of Meeran et al. (2013) was focused on predicting the sales of a new launched product, and Schwenke et al. (2012) simulated customers behavior to estimate the possible cart content and assess the changing in sales under different price conditions.

This paper contribution is twofold: (i) a switching dynamic model is proposed to perform both baseline and uplift forecasting; (ii) an imputation methodology, based on a similarity metric, is employed when promotional information are not present for a particular customer. The development of a switching model encompasses the benefits of both the last-like promotion and ADL models. The benefits of the proposed approach are the identification of a parametric model which is able to perform forecasts at any prediction horizon, encapsulating the different aspect of each promotion. Furthermore, when historical data are few, the ADL model structure can be too complex: the proposed method is shown to be reliable even with a low number of data, and domain specific KPIs (Key Performance Indexes) are introduced. The comparison of clients via a similarity measure can be used to produce visualizations and dashboards, enhancing the company understanding and decision making by revealing hidden groups and structures. In this view, the use of customer clustering, see Murray et al. (2015), is not new. However, in the aforementioned work clustering is used as a solution to build a lower number of models. Nowadays, provided that technological resources are available, it is not uncommon to deploy thousands of models into production, as described in Raeder et al. (2012). In this work, stores comparisons are instead used to recostruct missing promotional information at customer level.

The remainder of the paper is organized as follows. In Section 2, the business problem tackled in this work is stated. In Section 3, the steps involving the predictive model design and missing data imputation are presented. Section 4 highligths the main performance indexes used to assess the approach validity, and a comparison with known methods is performed. Section 5 is devoted to concluding remarks and future developments.

The whole analysis and method development will be supported by the use of experimental data taken from a real food company. For confidentiality reasons, data will be normalized and some details on the specific business will be omitted.

## 2. PROBLEM STATEMENT

The aim of this work is to evaluate the effect, that is, the sold quantity, generated by a different set of promotion types, for different products, for every customer of a food producer company. Forecasts are relative to the total delivered quantity in a working week. Correct predictions are beneficial for both the customers and producer. In the first case, there will be a more focused planning of the promotional calendar, while for the producer side, better inventory management and improved line production efficiency will be of direct conseguence.

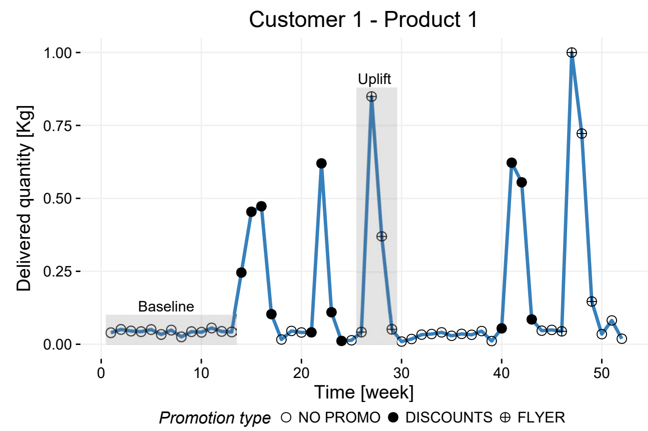


Fig. 1. Sales over a one year period. Baseline and uplifts phases are highlighted. The promotion of each week is represented by different symbols (see the legend).

The dataset employed in this work consists of 185 customers, 3 products (referred as Product 1, Product 2 and Product 3) and 8 promotion types (No Promo, Discount, Flyer, Collection Points, Flyer 1+1, Flyer discounted, Discount 1+1, Other). In each week, only a promotion can be present. A tipical behaviour of a sold quantity, normalized to protect sensible information, with and without promotional events, is depicted in Fig. 1. The promotion type "No Promo" indicates the absence of a promotion. The company needed to plan the working 2 weeks in advance, due to constraints on logistics, stocks and production management. The ultimate goal of a forecasting algorithm is to take optimal actions, which have a positive impact on the company economy, as discussed in the introduction. In this view, this work can be stated in a control system framework, where the focus is on the control component in Fig. 2.

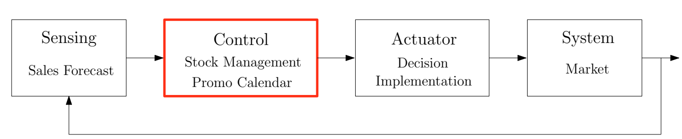


Fig. 2. Recasting of the business question as a control design problem: standing on the prediction of future requests, a promotion calendar can be modified or the inventory can be efficiently managed.

This boils down to the identification of a proper sales model (considered as the process under control), which will be used to infer future sales. This, accordingly, results in specific actions undertaken on production lines (the "Actuator" block). Once the forecasted quantity has been observed ("Sensing" block), corrective actions can be taken, both in the form of automatic model updating or human-in-the-loop interventions (the "Control" component). The feedback is generated by the fact that operational actions (select when and what promotion to apply, inventory replenishment decisions) affect the market. By looking at the market responses, new information about the effectiveness of certain promotion types and period of their application are gathered, improving the decision-making process. This framework permits an efficient, self-improving management of stocks and production operations. Similar control-oriented approaches applied to the business sector, such as inventory management, can be found in Azarskov et al. (2013); Abbou et al. (2015).

In order to evaluate the goodness of the proposed approach, three different KPIs were defined:

- **Integral Error during Promotion (IEP)**: to quantify the total error during promotional weeks
- **Mean Baseline Error (MEB)**: to check the error committed during non-promotional weeks
- **Maximum Uplift Error (MUE)**: to evaluate the maximum error committed in forecasting a single promotion effect

The proposed performance measures are considered more representative of the forecasting problem facets, with respect to standard evaluation metrics, such as the MAE (Mean Absolute Error, used in Trapero et al. (2014); Ali et al. (2009)), or the MAPE (Mean Average Percentage Error, see Yang et al. (2015)). In Tofallis (2015), the author showed the drawbacks of using such standard indicators, and proposed a metric based on the logarithm of the ratio between predicted and actual demand, called Log Accuracy Ratio (LAR). Nonetheless, MAE and MAPE metrics are reported for clarity, along with the LAR index. Information regarding customers sales were gathered from the company business intelligence software, which exports aggregates of selected delivered quantities. The program used for the computations is the open source R software stack.

## 3. SALES PREDICTION

This section describes the proposed approach to solve the forecast of promotional events' effect, illustrating first the parametric model identification procedure undertaken, and then the missing promotional information imputation strategy.

### 3.1 Covariate selection

For each customer-product couple, its relative dataset consisted in 104 weeks of historical data, with aggregated sales for each week. The data covered the 2014-2015 years. The data sampling time can therefore be considered as *one week*. The dependendet variable is the sold quantity, while chosen regressors consist of the promotion at each week (to evaluate the promotion effect on sales) and the period each week belong to (in order to obtain a stagionality effect). The variable "Period of the year" is computed by subdividing the weeks in a year into 6 periods (January-February, March-April, May-June, September-october, November-December). Given the assumption that previous sales level and promotions have an effect on the current delivered quantity, two additional covariates where computed from the existing ones. The new variables consist of the quantity delivered 2 weeks before the current time (in order to ensure the 2 weeks forecast horizon), and the promotion present in the week before the current one. Each week $t$ is then described by the following set of variables:

(1) Delivered quantity: $y(t)$ (*quantitative dependent variable*)
(2) Delivered quantity lagged of 2 timestamps: $y(t-2)$
(3) Promotion type: $p(t)$ (*categorical variable, 8 levels*)
(4) Promotion type lagged of 1 timestamp: $p(t-1)$
(5) Period of the year: $d(t)$ (*categorical variable, 6 levels*)

Given the low number of training points with respect to the number of regressors, a linear model structure was chosen to perform evaluations. The identification of a linear model is supported by the literature reviewed in Section 1: however, proposed methods like the ADL model, requires too many information which can not be available when a company has just started its business or does not have enough historical information. The proposed methodology consists in a switching model formulation, in order to estimate correctly both uplifts and baseline behaviours. Referring again to Fig. 1, it can be observed that the dependence of baseline periods on previous sales quantity is not evident, as opposite to uplift peaks. For this reason, the switching behaviour includes a static model which is more thrifty, being able to manage the low data available for the identification. During promotional events, a dynamic predictor is employed, which makes use of past information. During weeks with no promotions, the static model

$$\hat{y}(t) = \begin{cases} \alpha_d y(t-2) + \beta_d(p(t)) + \gamma_d(p(t-1)) + \mu_d(d(t)) + \\ +K[y(t-2) - \hat{y}(t-2)], \quad \text{if } p(t) \neq \text{NO PROMO} \\ \\ \beta_s(p(t)) + \mu_s(d(t)), \quad \text{if } p(t) = \text{NO PROMO} \end{cases} \tag{1}$$

estimates the baseline value. This switching strategy is made possible because the promotion calendar is known since the beginning of the year. The developed tool will help to improve the definition of the time and type of promotions for each week of the year, depending on the specific customer and product. Practical benefits of the proposed methodology are: i) independent estimation of baseline and uplift scenarios; ii) linear parametric structure, which makes simple the imputation of missing data; iii) effective even when the number of historical data is scarse.

The variable $\hat{y}(t)$ represents the predicted delivered quantity at time $t$, with $\alpha_d$, $\beta_d$, $\gamma_d$, $\mu_d$ the dynamic system estimated coefficients, and $\beta_s$, $\mu_s$ the static system estimated coefficients. The notation of the promotion coefficients $\beta_d$, $\gamma_d$, $\beta_s$ as function of the promotions at a certain time, indicates that various model with the same slope but different intercepts are being fitted by the linear model method. Variables $p(t)$ and $d(t)$ were coded as factor variables with defined levels, and $y(t)$ is a numerical quantity. The parameter $K$, chosen heuristically given the impossibility to perform proper cross-validation with the small data sample at disposal, controls the entity of the correction based on the previous 2-step prediction error. The model parameters are then estimated via least squares.

The model is trained incrementally each week, as more data are at disposal. This ensures that all available information are used and taken into account when performing forecasts. The initial training data set consists of the year 2014 data (52 observations), and predictions for the year 2015 in its entirety are performed. Then, the model is updated with data from the year 2015, and forecasts are recasted with the up to date sales information. As depicted in Fig. 3, is possible to observe how the model does a good job into estimating both uplifts and baseline values. It is to be noticed that the error on the promotion "Discount", represented by a filled circle, decreases as more data on this particular promotion type become available. The system is therefore obviously expected to perform even better with the arrival of new data.

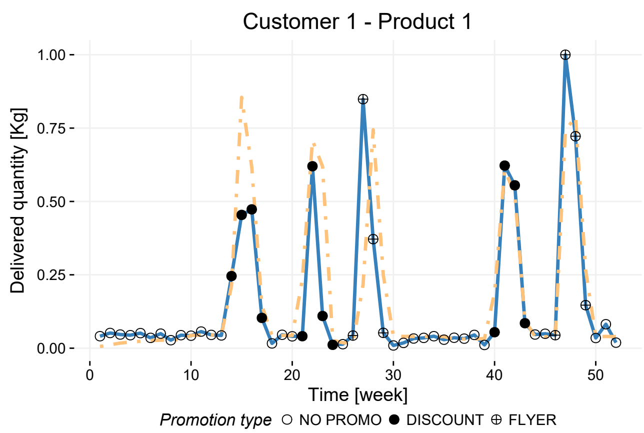


Fig. 3. Product 1 delivered quantity (continous blue line) and forecast (dot-dashed orange line). The large error on the first promotion uplift is due to the lack of information on *that* particular promotion in the training data (i.e., before week 13 of 2015).

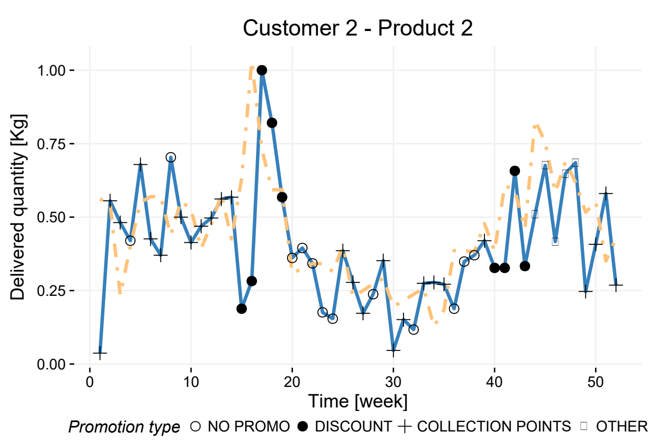


Fig. 4. Product 2 delivered quantity (continous blue line) and forecast (dot-dashed orange line).

The behaviour of the Product 2 for a specific customer 2, see Fig. 4, it is somewhat different with respect to the sales profile showed in Fig. 3. Although uplift peaks are still visibile, period with no promotion does not seem to differentiate much with respect to promotional weeks. Nevertheless, the forecasts are still able to capture the sales general trend. The plot in Fig. 5 reflects how the feedback correction acts, allowing a higher prediction at week 21 to compensate the underestimation of sales at week 19. From week 30, the Product 3 was no more sold in Customer 3 salespoints. The algorithm is able to cope with this sudden sales interruption. An interesting aspect

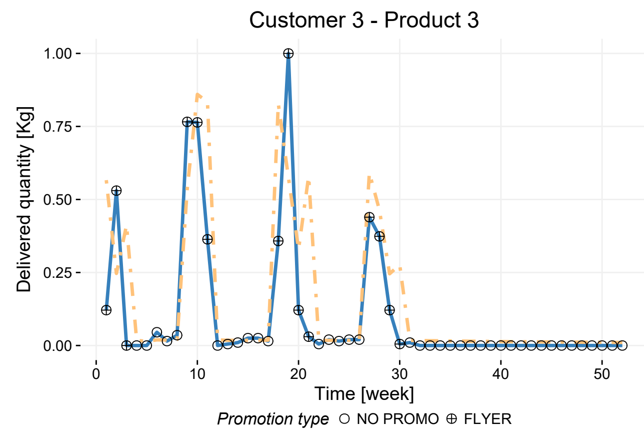


Fig. 5. Product 3 delivered quantity (continous blue line) and forecast (dot-dashed orange line).

is that the promotion "Fyler" in Fig. 3 is not present in the training data prior to week 28 of year 2015. To overcome this problem and being able to perform the required predictions, an imputation procedure based on a similarity measure between customers has been employed.

### *3.2 Missing data imputation*

In order to retrieve the missing promotional information, customers were compared by the euclidean distance measure, in the vector space defined by the following two features:

- **Mean baseline value**: the mean quantity sold during baseline periods
- **Mean uplift value**: the average increase in sales during promotional events. This feature is computed by taking the delivered quantity values at the beginning and at the end of the promotion, and their mean computed. Then, this value is subtracted from the maximum value of sold quantity during the promotion period. The procedure is repeated for each promotion of that particular client, giving raise to a number of indexes equal to the number of promotions of that client. Finally, their mean is computed and taken as feature

The adopted representation of clients in a 2-dimensional space permits a rapid and effective distance computations, given the number of stores (185 in this study). An higher dimensional space would have led to less significative comparisons due to the curse of dimensionality. As can be seen in Fig. 6 , the clients form one large cluster, with the upper-right customer considerable as another category. In order to impute missing promotional information for a client, its nearest customer (which has an evaluation of the missing promotion), is taken. Once the searched client is found, the model in Equation 1 is identified, and the

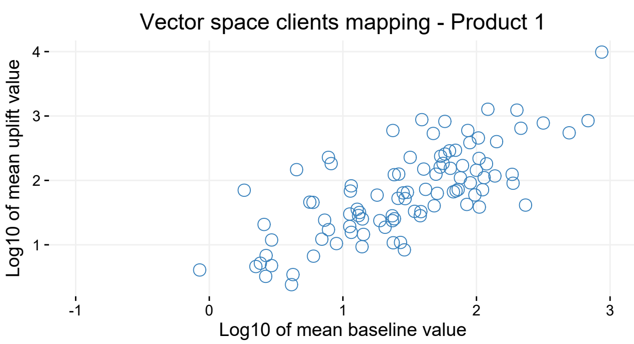


Fig. 6. Clustering induced on Product 1 clients.

coefficients corresponding to the missing promotions are used to evaluate the information not present. If the nearest customer does not have all required information (but only a subset of the missing information), the second nearest store is used, and so on. The results of this mapping can be used to interpret how a client is positioned into different market segments. The procedure has therefore been useful to develop a deeper understanding of the customer portfolio for the company.

## 4. PERFORMANCE ASSESSEMENT

In this section, the proposed model performances are compared with standard benchmark systems.

### *4.1 Last-like promotion benchmark model*

The first benchmark consists in the last-like promotion model, as implemented in Ali et al. (2009), with the minor modification of taking values two steps in the past. This method, also known as exponential smoothing with lift adjustment, contains a switching logic as the algorithm proposed in this paper. The philosopy behind the benchmark model is the following. If there is no promotion in the coming week, the forecast is the smoothed value of the non promotional weeks in the past; else, the last observed lift amount is added to the smoothed value to obtain the prediction:

$$\hat{y}_l(t) = \begin{cases} z(t) + L(t) & \text{if } p(t) \neq \text{NO PROMO} \\ z(t) & \text{if } p(t) = \text{NO PROMO} \end{cases} \tag{2}$$

and

$$z(t) = \begin{cases} z(t-2) & \text{if } p(t) \neq \text{NO PROMO} \\ (1-\tau)z(t-2) + \\ \tau y(t-2) & \text{if } p(t) = \text{NO PROMO} \end{cases} \tag{3}$$

with $\hat{y}_l(t)$ denoting the forecast at time $t$ of the last-like promotion model, $z(t)$ refers to the smoother number of items sold up to week $t$, based on non promotional weeks. The value of $\tau$ used is 0.2, based on the traditional default value for exponential smoothing. The lift amount $L(t)$ is computed as the difference of the actual sales at time $t$ and the smoothed non-promotion sales at the time of the most recent promotion, $L(t) = y(t-2) - z(t-2)$. The model behaves as a "persistency" one, replicating the uplift behaviour two weeks in the future.

### *4.2 ARMAX benchmark model*

The ARMAX (Autoregressive Moving Average with eXogenous inputs) model, see Ljung (1998), provides a general framework for modeling short-memory time series equipped with an external inputs $u(t)$:

$$A(z)y(t) = B(z)u(t) + C(z)e(t) \tag{4}$$

where $e(t)$ is a white noise error term, $z$ is the lag operator such that $z \cdot y(t) = y(t+1)$, and $A(z)$, $B(z)$, $C(z)$ are the polinomials of the autoregressive, exogenous and moving average part respectively. In this work, the most suitable model structure was found to be an ARMA(1,1) model, with indicators variable for period of the year and promotion type. The model was retrained each week, and a 2-step forecast computed.

### *4.3 Comparison of results*

A comparative overview of the proposed approach performance (on a client without missing promotion), with respect to the benchmarks described in the previous sections, is reported in Fig. 7. As can be seen, the last-like promotion model behaves like a persistence model, repeating the delivered quantity of two week before. The ARMAX model correctly detected the promotion periods, but it is not able to "reset" its prediction to the baseline value when the promotion ends. The proposed model is able to correctly detect the timings of a promotion, its effect, and rapidly switch to the baseline trend.

A quantitative comparison of the described methods is reported in Table 1 and Table 2, where the best results are highlighted in bold text. The proposed method is the best for 6 out of 7 KPIs. Major improvements lie in the baseline and uplifts value estimation. The ARMAX model instead obtains a better result on the integral error. Notice however that the proposed model is not designed to optimize such a measure but for a pointwise estimation.

Table 1. Comparison results: standard metrics

| **Method** | **MAE** | **MAPE** | **LAR** |
|---|---|---|---|
| Proposed | **0.171** | **195%** | **12.655** |
| Last-like | 0.418 | 249% | 35.404 |
| ARMAX(1,1) | 0.248 | 358% | 18.781 |

Table 2. Comparison results: proposed metrics

| **Method** | **IEP** | **MEB** | **MUE** |
|---|---|---|---|
| Proposed | 0.083 | **0.005** | **-0.186** |
| Last-like | -0.094 | 0.038 | -0.756 |
| ARMAX(1,1) | **0.019** | 0.056 | -0.514 |

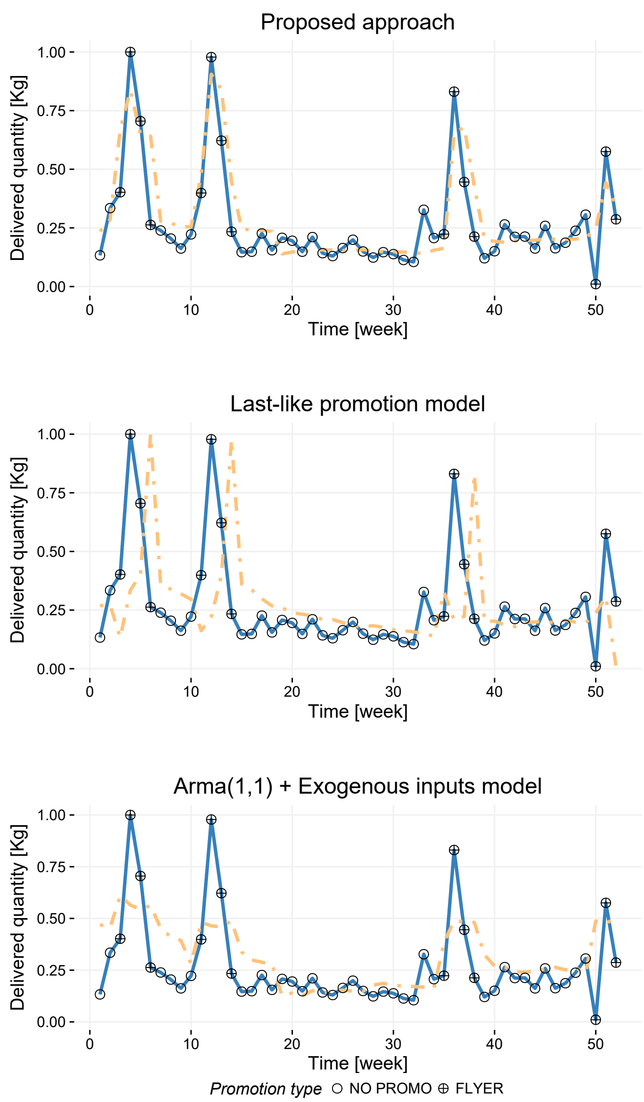


Fig. 7. Comparison of delivered quantity (continous blue line) and forecasts (dot-dashed orange line) with the proposed method and the benchmarks models.

## 5. CONCLUSIONS AND FUTURE DEVELOPMENTS

This paper presented an innovative approach to tackle the forecast of sales delivery in presence of promotional events. The proposed methodology consists in a switching linear structure, which is able to correctly predict both baseline and uplifts sale periods. The identified model has been shown to be particularly effective in presence of a low amount of training data for the identification stage. When no prior information on the effect of a promotion are available, an imputation procedure based on customer similarity has been proposed and employed with success.

Furthermore, a performance assessment and a comparison with state of the art methods has been carried out, introducing new and domain specific KPIs which better reflects the forecasts accuracy from different points of view.

Future research will be focused on the integration of the proposed modeling approach within a closed-loop business decision process.

## REFERENCES

Abbou, R., Moussaoui, C., and Loiseau, J.J. (2015). Effects of inventory control on bullwhip in logistic systems under demand and lead time uncertainties. *15th IFAC Symposium on Information Control Problems in Manufacturing (INCOM)*, 48(3), 266–271.

Abraham, M.M. and Lodish, L.M. (1987). Promoter: An automated promotion evaluation system. *Marketing Science*, 6(2), 101–123.

Ailawadi, K.L., Beauchamp, J.P., Donthu, N., Gauri, D.K., and Shankar, V. (2009). Communication and promotion decisions in retailing: a review and directions for future research. *Journal of retailing*, 85(1), 42–55.

Ali, Ö.G., Sayın, S., Van Woensel, T., and Fransoo, J. (2009). Sku demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10), 12340–12348.

Armstrong, J.S. (2001). *Principles of forecasting: a handbook for researchers and practitioners*, volume 30. Springer Science & Business Media.

Azarskov, V.N., Skurikhin, V.I., Zhiteckii, L.S., and Lypoi, R.O. (2013). Modern control theory applied to inventory control for a manufacturing system. *7th IFAC Conference on Manufacturing Modelling, Management, and Control (MIM)*, 46(9), 1200–1205.

Banek, M., Osrecki, D., Vranic, M., and Pintar, D. (2015). Outlier detection as the primary step for promotion planning in retail. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015 38th International Convention on*, 1405–1410. IEEE.

Blattberg, R.C. and Levin, A. (1987). Modelling the effectiveness and profitability of trade promotions. *Marketing Science*, 6(2), 124–146.

Brynjolfsson, E., Hitt, L.M., and Kim, H.H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN 1819486.*

Cooper, L.G., Baron, P., Levy, W., Swisher, M., and Gogos, P. (1999). Promocast: A new forecasting method for promotion planning. *Marketing Science*, 18(3), 301–316.

Holweg, M., Disney, S., Holmström, J., and Småros, J. (2005). Supply chain collaboration:: Making sense of the strategy continuum. *European management journal*, 23(2), 170–181.

Huang, T., Fildes, R., and Soopramanien, D. (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.

Ireland, R.K. and Crum, C. (2005). *Supply chain collaboration: How to implement CPFR and other best collaborative practices.* J. Ross Publishing.

Lee, W.Y., Goodwin, P., Fildes, R., Nikolopoulos, K., and Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, 23(3), 377–390.

Ljung, L. (1998). System identification. In *Signal Analysis and Prediction*, 163–173. Springer.

Meeran, S., Dyussekeneva, K., and Goodwin, P. (2013). Sales forecasting using combination of diffusion model and forecast market– an adaption of prediction/preference markets. *7th IFAC Conference on Manufacturing Modelling, Management, and Control (MIM)*, 46(9), 87–92.

Murray, P.W., Agard, B., and Barajas, M.A. (2015). Forecasting supply chain demand by clustering customers. *IFAC-PapersOnLine*, 48(3), 1834–1839.

Raeder, T., Stitelman, O., Dalessandro, B., Perlich, C., and Provost, F. (2012). Design principles of massive, robust prediction systems. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining*, 1357–1365. ACM.

Samson, D. and Singh, P.J. (2008). *Operations management: An integrated approach.* Cambridge University Press.

Schwenke, C., Ziegenbalg, J., Kabitzsch, K., and Vasyutynskyy, V. (2012). Simulation based forecast of supermarket sales. In *Proceedings of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA 2012)*, 1–8. IEEE.

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66(8), 1352–1362.

Trapero, J.R., Kourentzes, N., and Fildes, R. (2014). On the identification of sales forecasting models in the presence of promotions. *Journal of the Operational Research Society*, 66(2), 299–307.

Yang, D., Goh, G.S., Jiang, S., Zhang, A.N., and Akcan, O. (2015). Forecast upc-level fmcg demand, part ii: Hierarchical reconciliation. In *Big Data (Big Data), 2015 IEEE International Conference on*, 2113–2121. IEEE.