Forecasting with multivariate temporal aggregation: The case of promotional modelling

Nikolaos Kourentzes[a,*], Fotios Petropoulos[b]

[a]*Lancaster University Management School*
Department of Management Science, Lancaster, LA1 4YX, UK
[b]*Logistics and Operations Management Section*
Cardiff Business School, Cardiff University, UK

Abstract

Demand forecasting is central to decision making and operations in organisations. As the volume of forecasts increases, for example due to an increased product customisation that leads to more SKUs being traded, or a reduction in the length of the forecasting cycle, there is a pressing need for reliable automated forecasting. Conventionally, companies rely on a statistical baseline forecast that captures only past demand patterns, which is subsequently adjusted by human experts to incorporate additional information such as promotions. Although there is evidence that such process adds value to forecasting, it is questionable how much it can scale up, due to the human element. Instead, in the literature it has been proposed to enhance the baseline forecasts with external well-structured information, such as the promotional plan of the company, and let experts focus on the less structured information, thus reducing their workload and allowing them to focus where they can add most value. This change in forecasting support systems requires reliable multivariate forecasting models that can be automated, accurate and robust. This paper proposes an extension of the recently proposed Muliple Aggregation Prediction Algorithm (MAPA), which uses temporal aggregation to improve upon the established exponential smoothing family of methods. MAPA is attractive as it has been found to increase both the accuracy and robustness of exponential smoothing. The extended multivariate

*Correspondance: N Kourentzes, Department of Management Science, Lancaster University Management School, Lancaster, Lancashire, LA1 4YX, UK. Tel.: +44-1524-592911
Email address: n.kourentzes@lancaster.ac.uk (Nikolaos Kourentzes)

Preprint submitted to International Journal of Production Economics September 8, 2015

MAPA is evaluated against established benchmarks in modelling a number of heavily promoted products and is found to perform well in terms of forecast bias and accuracy. Furthermore, we demonstrate that modelling time series using multiple temporal aggregation levels makes the final forecast robust to model misspecification.

Keywords: Forecasting, temporal aggregation, MAPA, exponential smoothing, promotional modelling

---

1. Introduction

Demand forecasting is crucial for decision making and operations in organisations. As demand for large number of forecasts increases, for example due to the number of products companies trade, the reduced length of the forecasting cycle, or the increase in the number of item-location combinations as retail/logistic chains become larger and larger there is pressure to have reliable and accurate automated baseline forecasts. Typically companies rely on Forecasting Support Systems (FSS), which integrate univariate statistical baseline forecasts with managerial judgement to introduce additional external information in the forecasts (Fildes et al., 2006). Although the statistical element of such FSSs can be automated, the human element is resource intensive and often due to the increased workload, experts are not able to collect and account all relevant information into their forecasts. This is especially relevant for areas where numerous factors affect the demand, such as promotions and other marketing actions.

Past research has shown that judgmental adjustments indeed add value to the baseline statistical forecasts, but their performance is inconsistent (Fildes and Goodwin, 2007). Trapero et al. (2013) demonstrated that in the context of promotions it is desirable to include this additional information in statistical models that can be automated, therefore reducing the workload of human experts and allowing them to focus on incorporating less structured information in the forecasts. This research echoes a similar need often expressed by companies and practitioners. Therefore, there is demand to develop reliable multivariate statistical models that can be automated and address the requirements of large scale forecasting that organisations face nowadays.

There is extensive research in univariate forecasting methods, which are based on modelling the past time series structure and extrapolating it into

the future (Ord and Fildes, 2012). Well known methods include exponential smoothing and ARIMA, with the former being very widely used in practice, due to its simplicity, reliability and relatively good accuracy (Makridakis and Hibon, 2000; Gardner, 2006). The exponential smoothing family of methods is capable of modelling a wide variety of time series with or without trend and seasonality. With the incorporation of exponential smoothing in a state-space framework its statistical underpinnings were researched, resulting in an elegant and effective automatic model selection procedure (Hyndman et al., 2002, 2008). The basis of this model selection is to fit the various forms of exponential smoothing and choose the most appropriate based on a pre-selected information criteria, typically Akaike's (Hyndman et al., 2002; Billah et al., 2006). This approach has been implemented in various statistical software (Hyndman and Khandakar, 2008) and is widely regarded as a benchmark for automatic univariate forecasting that is at the core of FSSs.

More recently, further refinements in the automatic specification of exponential smoothing have appeared in the literature. From one hand, Kolassa (2011) argued that identifying a single model by using information criteria may not always perform well and investigated the performance of combining models via Akaike weights, instead of choosing a single one. He found this approach to be superior, resulting in more reliable and accurate forecasts. On the other hand, Kourentzes et al. (2014) looked at the combination of exponential smoothing models that are fitted across multiple temporally aggregated versions of the initial time series. They argued that their approach, named MAPA (Multiple Aggregation Prediction Algorithm) has advantages over conventional exponential smoothing modelling because different time series components are attenuated or strengthened at different temporal aggregation levels, resulting in a more holistic estimation of the time series structure and more accurate forecasts. Their approach builds on the extensive literature on the effects of temporal aggregation on forecasting (for recent examples see: Zotteri et al., 2005; Silvestrini and Veredas, 2008; Andrawis et al., 2011; Spithourakis et al., 2012; Rostami-Tabar et al., 2013).

However, these approaches are not able to make use of additional information such as promotions. Nonetheless, promotional modelling is crucial for many areas such as manufacturers of fast moving consumer goods and retailing. As argued above, automatic promotional forecasting is desirable. Regression type statistical models are often used to build promotional models (Fildes et al., 2008), which incorporate multiple exogenous marketing inputs. Such models are hard to automate and require substantial expertise

to maintain. Significant advances have taken place in promotional modelling at a brand level, involving sophisticated forms of regression (Cooper et al., 1999; Leeflang et al., 2002; Divakar et al., 2005). Yet, these models are not suited for SKU level forecasting that is relevant to the operations of organisations and alternative models have appeared in the literature making use of various regression type models and to a lesser extend ARIMA with external variables (Özden Gür Ali et al., 2009; Trapero et al., 2013, 2014; Huang et al., 2014). An apparent further candidate for this type of forecasting problems is exponential smoothing extended to include external variables (Hyndman et al., 2008; Athanasopoulos and Hyndman, 2008). Under this approach spate-space exponential smoothing can be enhanced to include additive exogenous effects following similar formulations as the aforementioned promotional models. Such models have not been explored in the literature, yet they are attractive due to the simplicity and good performance of the underlying method, as well as our good understanding on how to automate such models.

This paper investigates the use of multiple temporal aggregation to construct enhanced and automated exponential smoothing based promotional models. We extend the MAPA approach to include external variables, using a similar formulation to multivariate exponential smoothing. The motivation is to combine the simplicity and reliability of exponential smoothing with the estimation and robustness advantages of MAPA. We investigate the performance of the proposed method using a real case study of heavily promoted demand series of cider SKUs (Stock Keeping Units) of a popular brand in the UK. We use as benchmark the extended exponential smoothing that includes external promotional information, to demonstrate the advantages of using multiple temporal aggregation levels, and a regression based promotional model from the literature. We find that multivariate MAPA outperforms all benchmarks substantially, providing a useful candidate for a fully automatic promotional model. Furthermore, we find that exponential smoothing performs very well against regression based promotional models. We argue that one of the major advantages of the proposed method is its robustness to model misspecification and therefore its reliability for practical implementations.

The rest of the paper is organised as follows: section 2 describes MAPA and introduces our extension to model external variables; section 3 describes the case study that will be used to empirically evaluate the proposed method, while section 4 describes the experimental setup and the benchmarks used in

this research; section 5 presents the results, followed by a discussion on the benefits of temporal aggregation for promotional modelling and conclusions.

2. Methods

2.1. Multiple aggregation prediction algorithm

The Multiple Aggregation Prediction Algorithm (MAPA) was proposed by Kourentzes et al. (2014) to take advantage of the time series transformations that can be achieved by non-overlapping temporal aggregation. Temporally aggregating a time series can cause various of its components to become more or less prominent with direct effects on model identification and estimation. MAPA uses multiple temporal aggregation levels, allowing multiple views of the data to be considered during model building and subsequently combined in a final forecast.

MAPA can be seen as a three step procedure, where in the first step the original time series is aggregated in multiple aggregation levels using non-overlapping means of length k. The mean is used instead of the sum, as it retains the scale of the series across the various aggregation levels. Given a time series Y, with observations y_t and $t = 1, \ldots, n$, temporal aggregation can be performed as:

$$y_i^{[k]} = k^{-1} \sum_{t=1+(i-1)k}^{ik} y_t. \tag{1}$$

The temporally aggregated time series is noted with a superscript $[k]$ and has less observations than the original time series. For example for $k = 2$ the resulting series $Y^{[2]}$ will have half as many observations as the original time series. Note that the latter can be written under this notation as $Y^{[1]}$. Depending on the aggregation level k it may be that the division n/k has a non-zero remainder, in which case the $n - \lfloor n/k \rfloor k$ first observations of the time series are ignored in the construction of the aggregated one. The aggregation operator in Eq. (1) acts as a moving average and the resulting time series is smoother than the original one. High frequency components are progressively filtered as the aggregation level increases, essentially attenuating the seasonal and random component of time series, while allowing the low frequency trend and level components to dominate, capturing these better. Petropoulos and Kourentzes (2014b) suggested that aggregating up to time series of yearly time buckets it is sufficient, since all high frequency components will be

filtered by then, allowing to clearly see all low and high frequency elements of the series, although it is possible to consider even higher levels.

Subsequently, in the second step of MAPA a forecasting model is fitted at each aggregation level. Due to the aggregation operator it is expected that the original time series components will change. For fast moving consumer goods this means that seasonality may be present or trend easy to observe only some levels (Kourentzes et al., 2014), while for slow moving items the intermittency characteristics will change across the different aggregation levels, until the time series becomes non-intermittent (Petropoulos and Kourentzes, 2014a). Obviously, the underlying structure of the time series is constant, however due to the different sampling frequencies at the various aggregation levels, different elements of it become easier, more difficult or impossible to observe and estimate. Kourentzes et al. (2014) argued that this is a strength of the MAPA, as instead of selecting a single model, which may be wrongly identified, by repeating the process at each temporal aggregation level and combining the resulting models, potential problems due to errors in model selection and parametrisation are mitigated. However a new problem is introduced that results in the dampening of the estimated time series components. For example, let us assume that for a time series a seasonal model is estimated at one level, while a non-seasonal model is estimated at another. By combining the forecasts of these two levels the seasonal part is halved, assuming unweighted averaging is used. This is an undesirable property of forecast combination in the context of temporal aggregation, as it is expected that the time series components will not be present at all levels. To overcome this problem MAPA performs combination by time series components. The reader is referred to the discussion by Kourentzes et al. (2014) for more details.

Although in theory MAPA could use any forecasting method at each aggregation level, exponential smoothing is very suitable, as it separates a time series into level, trend and seasonal components during modelling. Exponential smoothing (ETS) models the level (l_t), trend (b_t) and seasonality (s_t) of a time series explicitly. These components are smoothed, and the level of smoothing is controlled by the smoothing parameters of ETS: α for the level, β for the trend and γ for the seasonal component. The smoothed components are then combined to give a forecast. Depending on the nature of the time series under consideration, these may interact in an additive or multiplicative way. Furthermore, the trend can be linear or damped, which is controlled by parameter ϕ. Table 1 provides the error correction forms of

Table 1: State space exponential smoothing equations for additive error

	Seasonal		
Trend	N	A	M
N	$\mu_t = l_{t-1}$ $l_t = l_{t-1} + \alpha\epsilon_t$	$\mu_t = l_{t-1} + s_{t-m}$ $l_t = l_{t-1} + \alpha\epsilon_t$ $s_t = s_{t-m} + \gamma\epsilon_t$	$\mu_t = l_{t-1}s_{t-m}$ $l_t = l_{t-1} + \alpha\epsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\epsilon_t/l_{t-1}$
A	$\mu_t = l_{t-1} + b_{t-1}$ $l_t = l_{t-1} + b_{t-1} + \alpha\epsilon_t$ $b_t = b_{t-1} + \beta\epsilon_t$	$\mu_t = l_{t-1} + b_{t-1} + s_{t-m}$ $l_t = l_{t-1} + b_{t-1} + \alpha\epsilon_t$ $b_t = b_{t-1} + \beta\epsilon_t$ $s_t = s_{t-m} + \gamma\epsilon_t$	$\mu_t = (l_{t-1} + b_{t-1})s_{t-m}$ $l_t = l_{t-1} + b_{t-1} + \alpha\epsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\epsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\epsilon_t/(l_{t-1} + b_{t-1})$
A_d	$\mu_t = l_{t-1} + \phi b_{t-1}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha\epsilon_t$ $b_t = \phi b_{t-1} + \beta\epsilon_t$	$\mu_t = l_{t-1} + \phi b_{t-1} + s_{t-m}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha\epsilon_t$ $b_t = \phi b_{t-1} + \beta\epsilon_t$ $s_t = s_{t-m} + \gamma\epsilon_t$	$\mu_t = (l_{t-1} + \phi b_{t-1})s_{t-m}$ $l_t = l_{t-1} + \phi b_{t-1} + \alpha\epsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\epsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\epsilon_t/(l_{t-1} + \phi b_{t-1})$
M	$\mu_t = l_{t-1}b_{t-1}$ $l_t = l_{t-1}b_{t-1} + \alpha\epsilon_t$ $b_t = b_{t-1} + \beta\epsilon_t/l_{t-1}$	$\mu_t = l_{t-1}b_{t-1} + s_{t-m}$ $l_t = l_{t-1}b_{t-1} + \alpha\epsilon_t$ $b_t = b_{t-1} + \beta\epsilon_t/l_{t-1}$ $s_t = s_{t-m} + \gamma\epsilon_t$	$\mu_t = l_{t-1}b_{t-1}s_{t-m}$ $l_t = l_{t-1}b_{t-1} + \alpha\epsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\epsilon_t/(s_{t-m}l_{t-1})$ $s_t = s_{t-m} + \gamma\epsilon_t/(l_{t-1}b_{t-1})$
M_d	$\mu_t = l_{t-1}b_{t-1}^{\phi}$ $l_t = l_{t-1}b_{t-1}^{\phi} + \alpha\epsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\epsilon_t/l_{t-1}$	$\mu_t = l_{t-1}b_{t-1}^{\phi} + s_{t-m}$ $l_t = l_{t-1}b_{t-1}^{\phi} + \alpha\epsilon_t$ $b_t = b_{t-1}^{\phi} + \beta\epsilon_t/l_{t-1}$ $s_t = s_{t-m} + \gamma\epsilon_t$	$\mu_t = l_{t-1}b_{t-1}^{\phi}s_{t-m}$ $l_t = l_{t-1}b_{t-1}^{\phi} + \alpha\epsilon_t/s_{t-m}$ $b_t = b_{t-1}^{\phi} + \beta\epsilon_t/(s_{t-m}l_{t-1})$ $s_t = s_{t-m} + \gamma\epsilon_t/(l_{t-1}b_{t-1}^{\phi})$

exponential smoothing with additive errors. The following notation is used: N for none, A for additive, A_d for additive damped, M for multiplicative and M_d for multiplicative damped. The forecast is denoted by μ_t and ϵ_t is the white noise error. Similar models exist for multiplicative error terms. To identify the correct form of ETS for each time series and temporal aggregation level the Akaike Information Criterion (AIC) is used, as it is suggested by Hyndman et al. (2002) for ETS modelling.

For MAPA we are interested in the last state vector $\boldsymbol{x}_i^{[k]}$ of ETS, which contains the updated values of each l_i, b_i and s_i: $\boldsymbol{x}_i^{[k]} = (l_i^{[k]}, b_i^{[k]}, s_i^{[k]}, s_{i-1}^{[k]}, \ldots, s_{i-m+1}^{[k]})'$. Using this information we can produce forecasts for any desirable horizon. Note that additive and multiplicative components will have different scale, as the later is expressed as a ratio of the level. This makes the combination by components difficult. To overcome this Kourentzes et al. (2014) proposed to first transform multiplicative components into additive using the formulae in table 2.

The additive translation of the components is only used for constructing

Table 2: Component prediction in the additive formulation

Trend	Seasonal		
	N	A	M
N	$l_{i+h} = l_i$	$l_{i+h} = l_i$ $s_{i-m+h} = s_{i-m+h}$	$l_{i+h} = l_i$ $s_{i-m+h} = (s_{i-m+h} - 1)l_{i+h}$
A	$l_{i+h} = l_i$ $b_{i+h} = hb_i$	$l_{i+h} = l_i$ $b_{i+h} = hb_i$ $s_{i-m+h} = s_{i-m+h}$	$l_{i+h} = l_i$ $b_{i+h} = hb_i$ $s_{i-m+h} = (s_{i-m+h} - 1)(l_{i+h} + b_{i+h})$
Ad	$l_{i+h} = l_i$ $b_{i+h} = \sum_{j=1}^{h} \phi^j b_i$	$l_{i+h} = l_i$ $b_{i+h} = \sum_{j=1}^{h} \phi^j b_i$ $s_{i-m+h} = s_{i-m+h}$	$l_{i+h} = l_i$ $b_{i+h} = \sum_{j=1}^{h} \phi^j b_i$ $s_{i-m+h} = (s_{i-m+h} - 1)(l_{i+h} + b_{i+h})$
M	$l_{i+h} = l_i$ $b_{i+h} = (b_i^h - 1)l_{i+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i^h - 1)l_{i+h}$ $s_{i-m+h} = s_{i-m+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i^h - 1)l_{i+h}$ $s_{i-m+h} = (s_{i-m+h} - 1)(l_{i+h} + b_{i+h})$
Md	$l_{i+h} = l_i$ $b_{i+h} = (b_i^{\sum_{j=1}^{h} \phi^j} - 1)l_{i+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i^{\sum_{j=1}^{h} \phi^j} - 1)l_{i+h}$ $s_{i-m+h} = s_{i-m+h}$	$l_{i+h} = l_i$ $b_{i+h} = (b_i^{\sum_{j=1}^{h} \phi^j} - 1)l_{i+h}$ $s_{i-m+h} = (s_{i-m+h} - 1)(l_{i+h} + b_{i+h})$

the out-of-sample component predictions that will be combined. Note that as these components are coming from different temporal aggregation levels, their length will be different. For example predicting at the monthly level a year ahead will result in twelve values, while in annual level will result in a single value. The translated component forecasts are returned to the original time domain using:

$$z_t = \sum_{j=1}^{k} \omega_j z_i^{[k]}, \tag{2}$$

where $z_i^{[k]}$ is the vector to be returned to the original time domain and $t = 1, 2, \ldots, n$ and $i = \lceil t/k \rceil$. Eq. (2) acts as a piecewise constant interpolation. The weights ω_j are equal to k^{-1}, resulting in an unweighted disaggregation scheme, which has been found to perform well (Nikolopoulos et al., 2011).

The last step of MAPA involves the combination of the components estimated across the different aggregation levels. Two combination methods were originally proposed: using unweighted mean and median, which were found to perform very similarly. In the case of the unweighted mean, each component is combined using:

$$\bar{l}_{t+h} = K^{-1} \sum_{k=1}^{K} l_{t+h}^{[k]}, \quad (3)$$

$$\bar{b}_{t+h} = K^{-1} \sum_{k=1}^{K} b_{t+h}^{[k]}, \quad (4)$$

$$\bar{s}_{t+h} = K'^{-1} \sum_{k=1}^{K'} s_{t+h}^{[k]}, \text{ if } (m/k) \in \mathbb{Z} \text{ and } k < m, \quad (5)$$

where K is the maximum aggregation level considered and K' is the number of aggregation levels where seasonality may be identified, i.e., when m/k results in an integer and $k < m$, as ETS is not capable of capturing fractional seasonality. The following example illustrates this: supposing a monthly sampled time series then $K' = 1, 2, 3, 4, 6$, i.e. seasonality estimated and combined only at monthly, bi-monthly, quarterly, four-month and semi-annual data. For trend, if at some aggregation level no trend is fitted, then it is assumed that for that level the value of trend is zero.

To produce the final forecast for h steps ahead, the forecast horizon of the original time series, the components can be simply added together, as they have been already translated into additive:

$$\hat{y}_{t+h}^{[1]} = \bar{l}_{t+h} + \bar{b}_{t+h} + \bar{s}_{t-m+h} \quad (6)$$

2.2. MAPA with exogenous variables

Here we will extend MAPA to include exogenous variables. Let X_j with observations $x_{j,t}$ be the j^{th} explanatory variable to be included in our model and $j = 1, \ldots, J$. The formulations in table 1 can be adjusted to include X_j as follows:

$$\tilde{\mu}_t = \mu_t + \sum_{j=1}^{J} d_{j,t}, \quad (7)$$

$$d_{j,t} = c_j x_{j,t},$$

where $d_{j,t}$ contain the effect of each X_j variable at time t and c_j is its coefficient. Coefficients c_j function in the same way as in a regression model, coding additive effects, while multiplicative effects can be captured through logarithmic transformation of the data. This formulation is similar to the standard ETS with regressor variables (Hyndman et al., 2008), with the only difference being that the effect of each variable is measured separately in $d_{j,t}$

allowing to directly incorporate it in the MAPA framework. Estimation of c_j is done simultaneously with the rest of the ETS states, μ_t. This can be done either by least squares, maximum likelihood estimation or other desirable cost functions.

At each temporal aggregation level k a separate $d_{j,i}^{[k]}$ is calculated, based on the estimated $c_j^{[k]}$ and temporally aggregated $X_j^{[k]}$. The resulting vectors are treated in the same way as the estimated time series components in the univariate case. First, they are translated into the original time domain using eq. (2). Then these are combined into a single effect across all aggregation levels for each variable X_j:

$$\bar{d}_{j,t+h} = K^{-1} \sum\nolimits_{k=1}^{K} d_{j,t+h}^{[k]}. \tag{8}$$

Finally, Eq. (6) that was used for the univariate forecast is adjusted to include the new multivariate effect estimations:

$$\hat{y}_{t+h}^{[1]} = \bar{l}_{t+h} + \bar{b}_{t+h} + \bar{s}_{t-m+h} + \sum\nolimits_{j=1}^{J} \bar{d}_{j,t+h}. \tag{9}$$

The parameters of the multivariate ETS at each temporal aggregation level will be optimised in the same way as the univariate ETS and the appropriate model form will be selected using AIC, as before. However, the temporal aggregation introduces one additional complexity for the multivariate models. As X_j are aggregated, they become smoother as implied by the aggregation Eq. (1). This changes the correlation between explanatory variables and may introduce multicollinearity at higher aggregation levels, if more than one variable is included in the model. As an illustrative example consider the case of two different promotions or special events that occur only once per month at a different day of the month and are coded using binary dummies. At a daily level these variables are not collinear, but at a monthly temporal aggregation level both variables become the same, equal to a vector of ones. Clearly, if both variables were included in Eq. (7) estimating coefficients c_j would not be possible. To avoid this it is desirable to transform the variables so that they become orthogonal. We can use principal components analysis to achieve this.

Principal component analysis generates a new set of variables, $\hat{X}_j$ with $j = 1, \ldots, J$, called principal components, which are linear combinations of the original variables. The weights of the linear combination are such that the resulting principal components are orthogonal to each other. Therefore,

the new variables $\hat{X}_j$ are no longer multicollinear and contain no redundant information (Jolliffe, 2002). These can now be used as inputs instead of the original X_j variables, overcoming the problems caused by temporally aggregating the explanatory variables. The principal components are constructed so that they are ordered in terms of variance, with the last components typically having very small variance. In practice we can omit theses, thus reducing the number of inputs to less than the original J. There are two commonly considered alternatives in choosing which components to retain. One can retain all components that are over a cut-off level in terms of variance. Alternatively, one can select to include only components that are significant in a regression context (Jolliffe, 1982). Here, for simplicity we use the first option, as conventional ETS parameter estimation does not typically provide standard errors of the estimated parameters that would allow the calculation of t-statistics. Note that it is still possible to obtain these by bootstrapping.

Therefore, by using principal components analysis we avoid the problem of multicollinearity of the inputs as the aggregation level increases and reduce the dimensionality of multivariate MAPA, making it less cumbersome to estimate.

Summarising, the extended MAPA works as follows. First the provided time series and promotions are temporally aggregated. At each aggregation level the data is processed as illustrated in the flowchart in Fig. 1. The promotional variables are first processed using principal components analysis and then incorporated in the exponential smoothing described by Eq. (7). From that the level, trend and seasonal components, as well as the promotional effect are extracted. The components are transformed to additive ones using the expressions in table 2. Then, together with the promotional part these are returned to their original frequency using Eq. (2). Estimates from all temporal aggregation levels are combined using Eqs. (3), (4), (5) and (8) for each level, trend, season and promotion components respectively. Finally these are combined in the final forecast using Eq. (9).

3. Case study

We empirically evaluate the performance of the multivariate MAPA by exploring its performance over benchmarks in predicting the sales of products under multiple promotions. Data from one of the leading cider brands have been collected from a UK manufacturer. These forecasts are useful for the manufacturer to support production and inventory planning decisions.

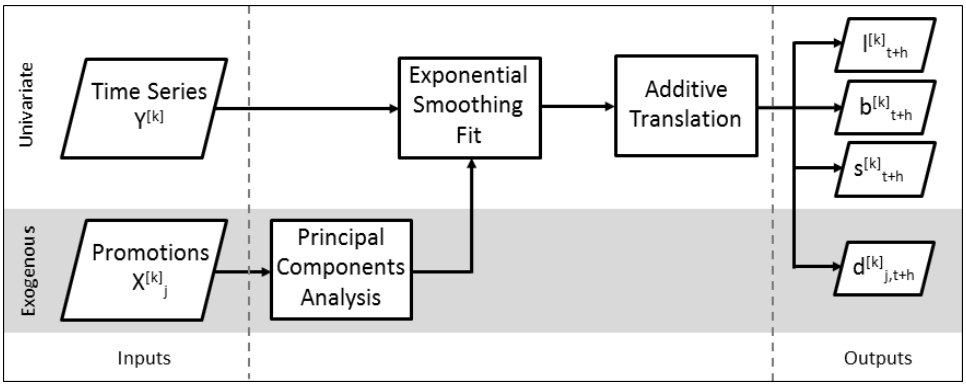

Figure 1: Flowchart of calculation steps for each temporal aggregation level of MAPA with exogenous variables.

Demand for 12 variants of the brand, including SKUs with different package sizes and flavours, has been collected for 104 weeks. The manufacturer sells the SKUs to multiple retailers who are offered different promotions. The timing of each promotion has been provided and was coded as binary dummy variables. Each SKU may be under up to 6 promotions at any time, accounting for the different offers to each retailer, with a varying degree of success. The promotions in this case study are known in advance, as the company has control of the promotional plan.

Table 3 provides the average descriptive statistics across SKUs. Looking at the difference between the measures of central tendency and the maximum we can observe the impact of promotions on sales, which is also reflected in the skewness of the sales. It can also be seen that these SKUs are heavily promoted, having on average 3.25 different promotions that are active for 61.78% of the sample. Note that all SKUs in the case study are fast moving.

Table 3: Average descriptive statistics across SKUs

Minimum	33.00
Mean	4038.81
Median	1959.67
Maximum	28151.75
Coefficient of variation	1.27
Skewness	2.56
Number of promotions	3.25
Periods under promotion	61.78%

As an example Fig. 2 provides the sales and the timing of the promotions to various retailers for a single SKU of the case study, which is representative of other SKUs in the dataset. Periods when at least one promotion takes place are highlighted. As Fig. 2 illustrates the SKU is under some promotion in almost every period. Note that for modelling the time series each retailer-level promotion is input separately so as not to assume that all have a similar effect.

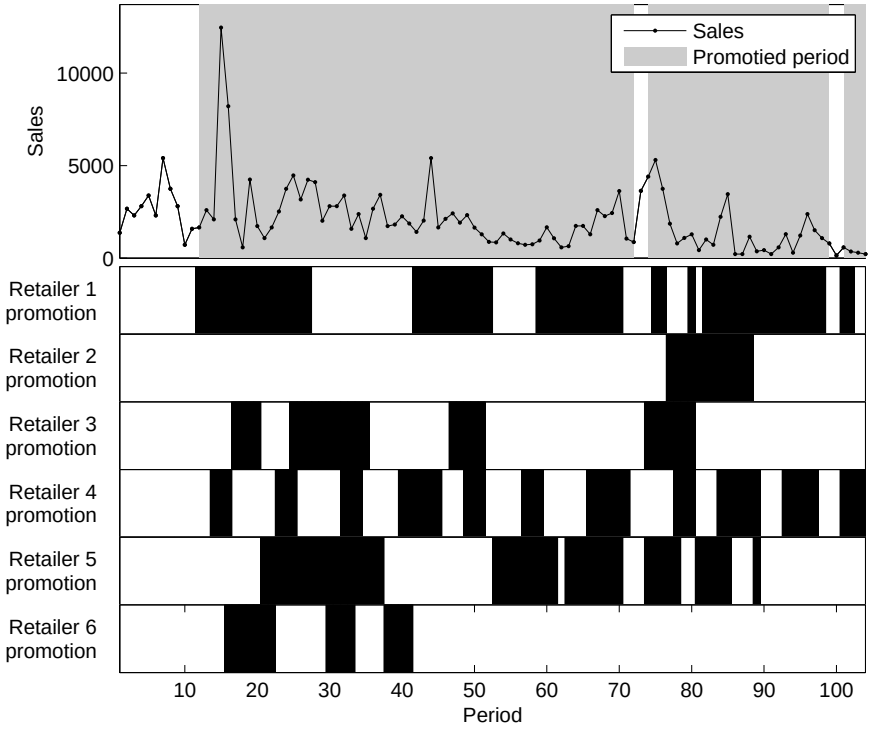

Figure 2: Sales and promotions of one SKU from the case study.

4. Empirical Evaluation

To evaluate the performance of the multivariate MAPA all SKUs available to us from the cider brand of our case study are used. For each time

series the last 18 weeks are withheld. This test set will be used to assess the out-of-sample forecasting performance of the method against established benchmarks for horizons t+4, t+8 and t+12, which are relevant for decision making for the manufacturer in our case study. We employ a rolling origin evaluation scheme to collect as many error measurements as possible for the three different forecast horizons in the test set. Forecasts are produced from each origin (week) of the out-of-sample period for the target forecast horizons and the performance is evaluated for each period (for more details on rolling origin evaluation see Tashman, 2000). The rest of the data is used for fitting the models. The methods are parametrised once in the in-sample period and used to produce all the forecasts in the test set.

We track the forecasting bias and accuracy for the weekly manufacturer sales of each SKU and horizon using the scaled Error (sE) and the scaled Absolute Error (sAE), which are defined as:

$$sE_t = \frac{y_t - f_t}{n^{-1}\sum_{i=1}^{n} y_i}, \tag{10}$$

$$sAE_t = \frac{|y_t - f_t|}{n^{-1}\sum_{i=1}^{n} y_i}, \tag{11}$$

where y_t and f_t are the actual and forecasted values at period t and the denominator is the mean of the time series. Both error metrics are scale independent and allow summarising the forecasting performance across the different time series. These errors are used instead of more common percentage metrics, such as the Mean Absolute Percentage Error, because the time series used in this study contain several periods of zero sales which makes the calculation of percentage metrics impossible. Furthermore, with traditional percentage errors periods with very low demand will have disproportionate impact. Both scaled metrics used here can be approximately interpreted as percentage forecast bias and error (Kolassa and Schütz, 2007). The error metrics are summarised across origins and time series by calculating the mean, resulting in the scaled Mean Error (sME) and scaled Mean Absolute Error (sMAE). For sME positive values imply under-forecasting and negative values imply over-forecasting.

The performance of multivariate MAPA is assessed using a number of benchmarks. First, a random walk forecast is used that will be referred to as *Naïve*. As the random walk is a very simple model that requires no parameter estimation, any more complex models should outperform it in order to

justify their additional complexity. Next, univariate *ETS* is used as a benchmark. Exponential smoothing is commonly used in business forecasting and has been found to be relatively accurate and reliable, both in practice and research (Gardner, 2006). The univariate *MAPA* introduced by Kourentzes et al. (2014) is also used as a benchmark, which has been shown to improve over the performance of *ETS*. Although both *ETS* and *MAPA* are not capable of modelling the available promotional information, they are useful benchmarks as they will permit us to evaluate the gains in performance achieved by their multivariate counterparts, if any. Due to limited estimation sample we consider temporal aggregation up to approximately the monthly level, $K = 4$.

Two multivariate benchmarks are used. In the literature there is a limited number of promotional models at SKU level (for examples see: Trapero et al., 2014; Huang et al., 2014). These differ from promotional models at brand level due to the different data structure and limitations: sales at SKU level are more disaggregate, having different time series components, increased noise and importantly limited data that prohibits fitting and using the substantially more complex and bigger in terms of variables brand level promotional models. Here we implement as a benchmark the model proposed by Trapero et al. (2014), which will be referred to as *Regression*, and was found to perform well. This is a regression based model that incorporates the following features: i) principal components analysis to reduce the dimensionality of model inputs and overcome to the multicollinearity of promotions that is often observed in practice; ii) modelling the promotion dynamics including potential lag effects; and iii) modelling the remaining time series dynamics that cannot be accounted for by the promotional activity, using ARMA components. The next multivariate benchmark is *ETS* with external regressors (Hyndman et al., 2008), referred to hereafter as *ETSx*. If the raw binary dummies are used as inputs the performance of this model is poor, due to the multicollinearity observed in the promotions. To overcome this we use principal components of the promotional dummies, following the suggestions by Trapero et al. (2014).

Finally, the multivariate *MAPA*, which will be referred to as *MAPAx*, is built as outlined in section 2.2. Similarly to *MAPA*, the maximum aggregation level considered for *MAPAx* is $K = 4$. Although principal components of the input variables are used in *MAPAx* due to the effects of temporal aggregation, at the same time this is beneficial in overcoming issues due to the multicollinearity of the promotional variables. In our experiments we found

that retaining only the first principal component at each aggregation level was adequate, substantially reducing the dimensionality of the model.

5. Results

Table 4 presents the results of the empirical evaluation across all available SKUs for the cider brand of the case study, in terms of sME and sMAE. The best performance for each error metric and horizon is highlighted in boldface. Values in parentheses represent medians across all SKUs, while the rest represent mean errors across SKUs. The last column in the table provides the mean rank of each method across SKUs and target forecast horizons for sME and sMAE. A method with rank of 1 is interpreted as being the best for every single case, while with rank of 6 it is always the worst.

Table 4: Mean (Median) forecasting bias (sME) and accuracy (sMAE) across SKUs

Method	t+4	t+8	t+12	Rank[†]
		sME		
Naïve	-0.139 (-0.022)	-0.194 (**+0.021**)	-0.282 (**-0.010**)	2.75
ETS	-0.249 (-0.204)	-0.287 (-0.328)	-0.374 (-0.371)	3.67
MAPA	-0.229 (-0.168)	-0.269 (-0.191)	-0.408 (-0.353)	3.83
Regression	-0.305 (-0.310)	-0.317 (-0.348)	-0.482 (-0.559)	4.42
ETSx	-0.214 (-0.112)	-0.171 (-0.147)	-0.250 (-0.220)	3.86
MAPAx	**-0.071** (**-0.021**)	**-0.048** (-0.029)	**-0.165** (-0.194)	**2.47**
MAPAx improvement over best benchmark	+48.9% (+4.5%) Naïve (Naïve)	+71.9% (-38.1%) ETSx (Naïve)	+34.0% (-94%) ETSx (Naïve)	-
		sMAE		
Naïve	0.743 (0.771)	0.818 (0.672)	0.704 (0.671)	3.75
ETS	0.704 (0.619)	0.774 (0.741)	0.701 (0.717)	3.86
MAPA	0.679 (0.611)	0.758 (0.679)	0.736 (0.727)	3.86
Regression	0.611 (0.579)	0.659 (0.642)	0.714 (0.682)	3.78
ETSx	0.642 (0.528)	0.627 (0.625)	0.543 (0.541)	3.06
MAPAx	**0.525** (**0.475**)	**0.521** (**0.447**)	**0.515** (**0.493**)	**2.69**
MAPAx improvement over best benchmark	+14.1% (+10.0%) Regr. (ETSx)	+16.9% (+28.48%) ETSx (ETSx)	+5.2% (+8.9%) ETSx (ETSx)	-

[†]Mean rank of method across horizons and SKUs. The method with the lowest reported rank performs best.

Overall, *MAPAx* is the best performer both in terms of average bias and error. It is interesting to evaluate the improvements achieved by extending

the models to use promotional information. To support the comparisons, Fig. 3 visualises the mean results presented in table 4. Focusing on the univariate *ETS* and *MAPA* the latter performs better for horizons t+4 and t+8 and the former for t+12. On average *MAPA* improves over *ETS*, in accordance to the findings by Kourentzes et al. (2014). This holds both in terms of forecast bias and error. Interestingly for long term forecasts, t+12, the *Naïve* has similar errors to both *ETS* and *MAPA*, attesting to the difficulty of producing accurate forecasts for the time series of our case study. When considering median errors the *Naïve* is more accurate for long term forecasts that both *ETS* and *MAPA*. In terms of bias the *Naïve* is always less biased.

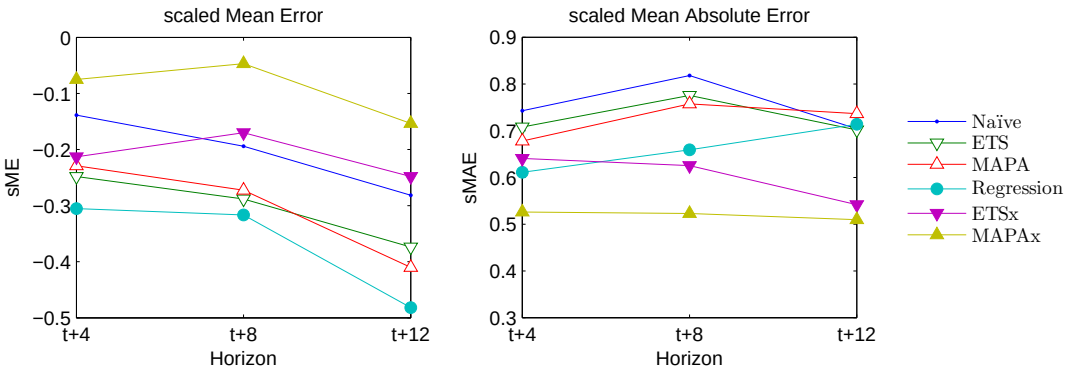

Figure 3: Mean forecast bias (sME) and error (sMAE) results.

When promotional information is included in the models their performance increases substantially. Starting from the benchmark *Regression* the forecast errors drop over the univariate models for horizons t+4 and t+8. For horizon t+12 the performance is again relatively poor, being similar to the *Naïve*. In terms of bias *Regression* is consistently the most biased. *ETSx* performs overall better than *Regression*, with the latter having lower errors only for the t+4 forecast horizons. In terms of median errors *ETSx* is always better than *Regression* and *Naïve*. It should be noted that in many ways *ETSx* incorporates several aspects of *Regression*, such as using principal components for the promotional information and capturing the time series dynamics. The primary difference between them is the way that the time series structure is identified and modelled, with *ETSx* being arguably simpler. Furthermore *ETSx* has substantial performance improvements over

its univariate counterpart, demonstrating the benefits of including the promotional information.

MAPAx exhibits the biggest improvement over its univariate counterpart. The observed improvements demonstrate again the benefit of including promotional information in the models. Considering mean sME across SKUs *MAPAx* gives the least biased predictions, with substantial differences over *ETSx* for all forecast horizons. However, when medians are considered *MAPAx* is second after the *Naïve* for longer horizons (t+8 and t+12). Nonetheless, it still exhibits substantial improvements over all other methods and in particular *Regression* and *ETSx* that are capturing the promotional information. In terms of accuracy *MAPAx* has lower errors than both multivariate benchmarks, considering either mean or median errors across SKUs. Overall, considering the mean errors of the best performing benchmark for each horizon, *MAPAx* is about 51.6% less biased and has about 12.0% lower forecast errors.

Focusing on the mean ranks provided in table 4, *MAPAx* achieves the best ranking for both sME and sMAE, demonstrating its consistent performance. The value of the promotional inputs is highlighted in the mean ranks of sMAE, where *Regression* and *ETSx* rank better than the univariate benchmarks. This demonstrates that the promotional inputs are useful for improving forecasting accuracy. Note that *Naïve* performs better than the univariate *ETS* and *MAPA* providing evidence of the difficulty of producing accurate baseline forecasts for the time series of the case study.

In many ways the relative performance of *MAPAx* in comparison to *ETSx* replicates the pattern between the univariate *MAPA* and *ETS*. Using multiple temporal aggregation consistently results in better performance over conventionally modelled exponential smoothing.

Therefore the superior performance of *MAPAx* is a result of the combination of the quality of the forecasting method and the quality of information available to it. These results were found to be consistent using other error metrics, such as scaled Mean Squared Error.

6. Discussion

Considering the conventional *ETS* if there are strong promotional effects, as it is true in our case study, the parameter estimates and even the selected model, as it is conditional on the estimated parameters, may be biased. By introducing the promotional information in the *ETSx* model this effect is

mitigated. However, there is still uncertainty in the parameter identification and model selection, due to available sample and sampling frequency issues (Kourentzes et al., 2014) or the inherent limitations of information criteria for model selection (Kolassa, 2011). The original *MAPA* was developed with the motivation of addressing the later issues. The time series is modelled at multiple temporal aggregation levels, thus at each level filtering the higher frequency components of the time series, allowing to estimate lower ones appropriately. Combining the estimates across the different aggregation levels results in robust final forecasts, as there is little reliance on a single model or a single view –aggregation level– of the time series, gaining the advantages of model combination. Nonetheless, similar to *ETS* and *ETSx*, the various models estimated under *MAPA* for the different aggregation levels will be biased if no promotional information is provided under the presence of strong effects. *MAPAx* address this by taking advantage of the additional information.

The effect of including promotional information at low levels of aggregations is apparent, as at this level their effect will be stronger. However, at higher aggregation levels the size of the effect of promotions at each period becomes smaller and one could expect that it is no longer as important. Eq. (1) shows that aggregation acts as a moving average, therefore although the effect per period will be smaller, the promotion now is expanded to neighbouring periods. This results again in an important overall effect, which unless modelled explicitly it is bound to bias parameter estimates and potentially even the selection of the model for each aggregation level.

Let us consider the example of a simulated sales series with promotions. Fig. 4 plots the sales series at various temporal aggregation levels. The promoted periods are noted with black bars at the lower part of the plots. Furthermore, for comparison, the simulated sales as if there were no promotions are plotted with a dotted line. Observe that the sales do not contain any trend or seasonality and therefore the only non-promotional structure is the level. Single exponential smoothing would be appropriate to produce forecasts if there were no promotions, where the single smoothing parameter α captures the dynamics of the time series levels. To illustrate the effect of including the promotional information on the parameter estimates table 5 provides the estimated smoothing parameter α at each aggregation level. The aim of this example is to illustrate the effect that promotional information has at various aggregation levels on the estimation of the level component.

The first column, *ETS - sales without promotions*, lists the fitted parame-

ters for the simulated sales series without promotional effects, corresponding to the dotted line in Fig. 4. The second column, *ETS - sales with promotions*, lists the fitted α parameters when the sales series includes the peaks due to promotions. Note that in all cases the parameters are substantially different, demonstrating the impact of the promotions not captured on model fit. Now the level of the time series is modelled wrongly and the accuracy of the forecasts is expected to be poor. Interestingly this is true even for high aggregation levels that the promotional uplift is seemingly small.

The last column of table 5, *ETSx*, lists the α parameters of *ETSx* fits that model promotions as an additional input. Although the parameters are not identical to those of the first column, they are much closer demonstrating the advantage of including such information when available, even when its effect is relatively smoothed due to the temporal aggregation. Now the level dynamics are captured more accurately and the resulting forecasts are expected to perform better.

It is also interesting to note that there may be cases, as is for aggregation level 6 in this example, where the smoothing parameter is apparently misestimated. In these cases the forecasts of *ETSx* will be of poor quality, while the ones of *MAPAx* that combine the estimates from multiple aggregation levels will be better. A similar observation can be made with regards to the fitted model at each aggregation level as argued by Kourentzes et al. (2014). Potentially models fitted at some aggregation levels may be misspecified in their form. By using multiple temporal aggregation, as the outputs of the various models at the different aggregation levels are combined, we do not rely on a single one, which might have been misspecified. This is a useful property for practical implementations of *MAPAx*, as it makes it robust against misspecification at some aggregation levels and crucially at the original time series, which is the only view of the data conventional time series modelling focuses on. Therefore *MAPAx* provides a reliable automatic forecasting procedure that includes external variables.

Table 5: Smoothing parameter α for simulated sales example

Aggregation Level	ETS - sales without promotions	ETS - sales with promotions	ETSx
1	0.046	0.093	0.054
2	0.137	0.215	0.128
4	0.252	0.146	0.198
6	0.230	1.000	0.000

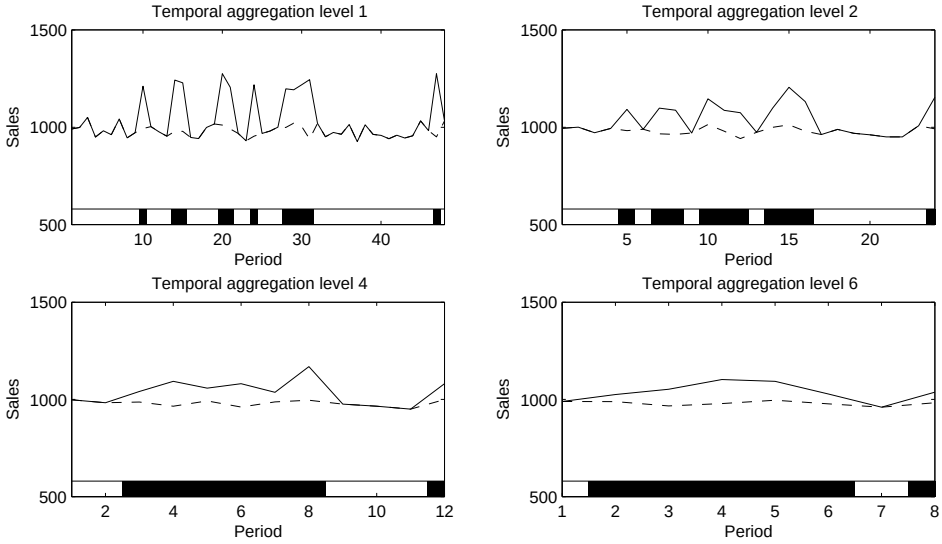

Figure 4: Sales with promotions at temporal aggregation levels 1, 2, 4 and 6. Periods under promotion are noted with black bars at the lower part of the plots.

MAPAx is useful for practice as it incorporates explanatory variables in an automated way, such as promotions, and provides reliable and accurate predictions. This makes it useful for supply chain forecasting, where typically a large number of SKUs need to be forecasted for inventory and planning purposes. Therefore it is interesting to consider the implications of using *MAPAx* for such cases. Stock calculations are typically based on the following formula: expected demand over lead time plus demand uncertainty over lead time. The first quantity is essentially the expected value of the forecast, which ideally should have a forecast bias of zero, otherwise the expected value of the forecast does not match the expected value of the demand. The second quantity, which is essentially the safety stock, is a pre-set percentile of the distribution of forecast error size, which is often approximated as the mean squared error of the forecast multiplied by some factor to account for the target service level and the lead time. Therefore, a good forecast for such purposes should have small bias and magnitude of forecast errors. Table 4 provided evidence of the superior performance of *MAPAx* in our case study, both in terms of forecast bias and error, demonstrating that it has desirable behaviour and outperforms the various benchmarks in both dimensions.

7. Conclusions

This paper extended the univariate Multiple Aggregation Prediction Algorithm that was recently proposed in the literature, and has been shown to have good performance both for fast and slow moving items, to the multivariate case. To demonstrate the performance and the efficacy of the proposed formulation we investigated the usage of the *MAPAx* to model the demand of SKUs of a popular cider brand in the UK, including promotional information.

MAPAx was found to outperform all benchmarks, which included a recently proposed in the literature SKU-level promotional model and exponential smoothing with regressor inputs, appropriately preprocessed. In particular, the main differences between *ETSx* and *MAPAx* is the use of multiple temporal aggregation levels, which provides the latter approach its superior performance and also makes it robust against model misspecification. The overall better performance of *MAPAx* over its exponential smoothing counterpart follows similar findings for the univariate case in the literature, providing evidence of the merits of this alternative approach to forecasting time series, based on modelling time series at multiple temporal aggregation levels.

In the discussion we attempted to highlight the implications of using *MAPAx* for baseline forecasting in a supply chain context. Future research should explore in detail the inventory implications of using *MAPAx* when external variables are available and important for capturing the demand behaviour. Another aspect of using *MAPAx* in a supply chain context that warrants further research is the interaction of human experts with the statistical forecast. As forecasting methods become more complex, here to introduce promotional information at SKU level, their transparency to experts is reduced, complicating the adjustment process.

Acknowledgement

We would like to thank the anonymous reviewers for their comments that helped us improve the paper. Any remaining errors are the responsibility of the authors.

References

Andrawis, R. R., Atiya, A. F., El-Shishiny, H., 2011. Combination of long term and short term forecasts, with application to tourism demand forecasting. International Journal of Forecasting 27 (3), 870 – 886.

Athanasopoulos, G., Hyndman, R. J., 2008. Modelling and forecasting australian domestic tourism. Tourism Management 29 (1), 19–31.

Billah, B., King, M. L., Snyder, R. D., Koehler, A. B., 2006. Exponential smoothing model selection for forecasting. International Journal of Forecasting 22 (2), 239–247.

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

Divakar, S., Ratchford, B. T., Shankar, V., 2005. CHAN4CAST: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. Marketing Science 24, 334–350.

Fildes, R., Goodwin, P., 2007. Against your better judgment? How organizations can improve their use of management judgment in forecasting. Interfaces 37 (6), 570–576.

Fildes, R., Goodwin, P., Lawrence, M., 2006. The design features of forecasting support systems and their effectiveness. Decision Support Systems 42 (1), 351–361.

Fildes, R., Nikolopoulos, K., Crone, S., Syntetos, A., 2008. Forecasting and operational research: a review. Journal of the Operational Research Society 59 (9), 1150–1172.

Gardner, E. S., 2006. Exponential smoothing: The state of the art - part II. International Journal of Forecasting 22 (4), 637–666.

Huang, T., Fildes, R., 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.

Hyndman, R. J., Khandakar, Y., 7 2008. Automatic time series forecasting: The forecast package for R. Journal of Statistical Software 27 (3), 1–22.

Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2008. Forecasting with Exponential Smoothing: The State Space Approach. Springer Verlag, Berlin.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. International Journal of Forecasting 18 (3), 439–454.

Jolliffe, I. T., 1982. A note of the use of Principal Component in regression. Applied Statistics 31, 300–303.

Jolliffe, I. T., 2002. Principal Component Analysis., 2nd Edition. Springer series in statistics. Springer, New York.

Kolassa, S., 2011. Combining exponential smoothing forecasts using akaike weights. International Journal of Forecasting 27 (2), 238–251.

Kolassa, S., Schütz, W., 2007. Advantages of the MAD/Mean Ratio over the MAPE. Foresight: The International Journal of Applied Forecasting (6), 40–43.

Kourentzes, N., Petropoulos, F., Trapero, J. R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. International Journal of Forecasting 30 (2), 291–302.

Leeflang, P. S., van Heerde, H. J., Wittink, D., 2002. How promotions work: SCAN*PRO-based evolutionay model building. Schmalenbach Business Review 54, 198–220.

Makridakis, S., Hibon, M., 2000. The M3-competition: results, conclusions and implications. International Journal of Forecasting 16 (4), 451–476.

Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., Assimakopoulos, V., 2011. An aggregate–disaggregate intermittent demand approach (adida) to forecasting: an empirical proposition and analysis. Journal of the Operational Research Society 62 (3), 544–554.

Ord, J. K., Fildes, R., 2012. Principles of Business Forecasting, 1st Edition. South-Western Cengage Learning, Mason, Ohio.

Özden Gür Ali, Sayin, S., van Woensel, T., Fransoo, J., 2009. SKU demand forecasting in the presence of promotions. Expert Systems with Applications 36 (10), 12340 – 12348.

Petropoulos, F., Kourentzes, N., 2014a. Forecast combinations for intermittent demand. Journal of Operational Research Society.

Petropoulos, F., Kourentzes, N., 2014b. Improving forecasting via multiple temporal aggregation. Foresight: The International Journal of Applied Forecasting 32, 12–17.

Rostami-Tabar, B., Babai, M. Z., Syntetos, A., Ducq, Y., 2013. Demand forecasting by temporal aggregation. Naval Research Logistics (NRL) 60 (6), 479–498.

Silvestrini, A., Veredas, D., 2008. Temporal aggregation of univariate and multivariate time series models: A survey. Journal of Economic Surveys 22 (3), 458 – 497.

Spithourakis, G., Petropoulos, F., Nikolopoulos, K., Assimakopoulos, V., 2012. A systemic view of ADIDA framework. IMA Management Mathematics forthcoming.

Tashman, L. J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. International Journal of Forecasting 16 (4), 437–450.

Trapero, J. R., Kourentzes, N., Fildes, R., 2014. On the identification of sales forecasting models in the presence of promotions. Journal of the Operational Research Society.

Trapero, J. R., Pedregal, D. J., Fildes, R., Kourentzes, N., 2013. Analysis of judgmental adjustments in the presence of promotions. International Journal of Forecasting 29 (2), 234–243.

Zotteri, G., Kalchschmidt, M., Caniato, F., 2005. The impact of aggregation level on forecasting performance. International Journal of Production Economics 93 - 94, 479 – 491.