Journal of Business Research 78 (2017) 1–9

Contents lists available at ScienceDirect

Journal of Business Research

journal homepage: www.elsevier.com/locate/jbusres

Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels?

Nikolaos Kourentzes[a,*], Bahman Rostami-Tabar[b], Devon K. Barrow[c]

[a] Lancaster University Management School Department of Management Science, Lancaster LA1 4YX, UK
[b] Cardiff Business School, Cardiff University Cardiff CF10 3EU, United Kingdom
[c] School of Strategy and Leadership, Faculty of Business and Law Coventry University, Coventry, West Midlands CV1 5FB, UK

ARTICLE INFO

Keywords:
Forecasting
Demand planning
Temporal aggregation
Model selection
Exponential smoothing
MAPA

ABSTRACT

Recent advances have demonstrated the benefits of temporal aggregation for demand forecasting, including increased accuracy, improved stock control and reduced modelling uncertainty. With temporal aggregation a series is transformed, strengthening or attenuating different elements and thereby enabling better identification of the time series structure. Two different schools of thought have emerged. The first focuses on identifying a single optimal temporal aggregation level at which a forecasting model maximises its accuracy. In contrast, the second approach fits multiple models at multiple levels, each capable of capturing different features of the data. Both approaches have their merits, but so far they have been investigated in isolation. We compare and contrast them from a theoretical and an empirical perspective, discussing the merits of each, comparing the realised accuracy gains under different experimental setups, as well as the implications for business practice. We provide suggestions when to use each for maximising demand forecasting gains.

1. Introduction

Demand forecasting plays a crucial role in the operations of modern organisations (Fildes, Nikolopoulos, Crone, & Syntetos, 2008; Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016). It supports a variety of business decisions, from operational, to tactical, to strategic level, such as capacity planning (Miyaoka & Hausman, 2008), resource planning (Barrow, 2016; Jalal, Hosseini, & Karlsson, 2016), advertising and promotional planning (Trapero, Kourentzes, & Fildes, 2014; Ma, Fildes, & Huang, 2016), demand planning (Trapero, Kourentzes, & Fildes, 2012; Syntetos, Babai, & Gardner, 2015), analysing competition effects (Merino & Ramirez-Nafarrate, 2016), tactical production planning (Sagaert, Aghezzaf, Kourentzes, & Desmet, 2017), among others. Accordingly, practitioners need to define the forecast objective in terms of forecast horizon and time bucket (e.g. daily, weekly, monthly, etc.), so as to support the appropriate decisions.

An important assumption is that the level of required forecasting matches the level of available collected data. However, often this is not true. For example, in many organizations, managers from several departments are involved in forecast generation and adjustment, that supports decisions for production, inventory management, logistics, procurement, and others (Lapide, 2004); with each function having different decision horizons. For example, budget forecasts are not required at the, typically, weekly resolution of inventory management, and refer to much longer horizons than the latter.

As a remedy the original data series can be aggregated over time (temporal aggregation, TA) to align the decision parameters with the forecast modelling, or alternatively disaggregated. Recently there has been a resurgence in researching TA for forecasting. In the past the research had mostly focused in modelling macroeconomic time series, but current work has demonstrated its usefulness for forecasting business time series, and in particular for the purpose of demand forecasting to support decision-making in operations management (Babai, Ali, & Nikolopoulos, 2012; Kourentzes & Petropoulos, 2016a; Boylan & Babai, 2016). Using TA a time series is modelled at a pre-specified aggregation level, instead of its original sampling frequency. Forecasts are then created at the aggregate level, which may be disaggregated to the original frequency, if so needed. The motivation for using TA is that it smooths the original series, removing noise and even some of its component, simplifying the generation of forecasts, which is desirable in itself (Green & Armstrong, 2015). The exact effects depend on the selected aggregation level, a critical consideration for the effectiveness of TA.

To this end, the econometric literature has explored the effect of TA, mainly on AutoRegressive Integrated Moving Average (ARIMA) processes (Silvestrini & Veredas, 2008), providing some evidence of the

* Corresponding author at: Department of Management Science, Lancaster University Management School, Lancaster, Lancashire LA1 4YX, UK.
E-mail address: n.kourentzes@lancaster.ac.uk (N. Kourentzes).

http://dx.doi.org/10.1016/j.jbusres.2017.04.016
Received 14 October 2016; Received in revised form 21 April 2017; Accepted 22 April 2017
0148-2963/

benefits and caveats of the practice, while more recent forecasting research has helped identify analytically the optimal aggregation level for a small number of processes, under specific modelling conditions (Rostami-Tabar, Babai, Syntetos, & Ducq, 2013, 2014). Nonetheless, general guidelines for how to best select the aggregation level do not exist, and this introduces substantial uncertainty in the modelling process. This has lead Kourentzes, Petropoulos, and Trapero (2014) to propose using multiple TA levels instead of a single one. In this case, modelling happens at multiple levels and the output is a combined forecast.

Therefore, although there is a strong theoretical and empirical evidence that TA can be beneficial to forecasting, there is no consensus as to how best perform it. The two alternative schools of thought recommend from the one hand to use a single optimal TA level, and from the other hand to use multiple levels, since the identification of a single level is problematic. The aim of this paper is threefold: (i) we contrast the two approaches both from a theoretical and empirical perspective; (ii) we benchmark these against heuristic based alternatives; and (iii) provide additional evidence of the usefulness of TA for demand forecasting over traditional time series modelling, at the original sampling frequency.

We find that overall TA is beneficial for demand forecasting over conventional time series modelling. Each school of thought offers different advantages and has different limitations. The main limitation of identifying an optimal single aggregation level is that it assumes knowledge of the demand process at both the original and the aggregate level, with the obvious implications for practice. On the other hand, using multiple levels is particularly robust to model uncertainty and is found to provide accuracy improvements for wide number of cases. However, the forecast is suboptimal by design in the strict sense of mean squared error fit. Finally, we translate these findings to implications for business forecasting practice.

The rest of the paper is organised as follows: Section 2 provides an overview of the use and developments of TA in demand forecasting and Section 3 describes the two alternative approaches in using TA. Section 4 describes the datasets used and the setup of our evaluation, while Section 5 presents the results, followed by concluding remarks in Section 6.

2. Temporal aggregation in business forecasting

Non-overlapping TA can be seen as a filter of high-frequency components of the time series. As we aggregate, low frequency components will dominate and depending on the level of aggregation higher frequency components will become weaker or vanish altogether. For example, consider a monthly seasonal time series that is aggregated to an annual series. The high frequency seasonal component is filtered, while the observed variance of the time series will be mostly due to the trend/cycle component.

In the econometric literature TA has been researched for several decades and the focus has mainly been on its effects on ARIMA processes. The key theoretical results can be summarised as follows: (i) TA reduces the number of available observations; hence causing loss of estimation efficiency; (ii) the dynamics of the underlying ARIMA process become more complicated, mainly due to the moving average component; and (iii) the identifiable ARIMA converge to elatively simple IMA processes, often IMA(1,1) (Wei, 1978; Rossana & Seater, 1995). The literature provides evidence of accuracy gains of forecasting directly using temporally aggregated data, rather than aggregating forecast from disaggregate series (Silvestrini & Veredas, 2008).

2.1. Temporal aggregation at a single level

More recently there has been substantial research on TA for business forecasting and supply chain management. Nikolopoulos, Syntetos, Boylan, Petropoulos, and Assimakopoulos (2011) recommend using TA for modelling and forecasting intermittent time series in a supply chain context. Their main motivation is to avoid modelling the intermittency at the sampling frequency directly and instead model the series with conventional forecasting methods, once the intermittency has been reduced substantially. They demonstrate that on average TA provides accuracy improvements. This finding has been validated several times in the context of intermittent demand forecasting (Babai et al., 2012; Petropoulos & Kourentzes, 2014a). It is important to note that Nikolopoulos et al. (2011) do not provide a conclusive solution with regards to the identification of the appropriate TA level. Instead, they recommend a heuristic that is meaningful for inventory management: aggregate to the level that corresponds to the lead time plus review period. Petropoulos, Kourentzes, and Nikolopoulos (2016) demonstrate that some intermittent demand forecasting methods, such as Croston's method, can be interpreted as special cases of TA and propose various alternative setups of TA, which in turn can reduce the variability of the non-zero demand or the inter-demand intervals and demonstrate benefits for forecast accuracy.

Spithourakis, Petropoulos, Babai, Nikolopoulos, and Assimakopoulos (2011) extended the work by Nikolopoulos et al. (2011) to fast moving demand data, validating that TA leads to forecast accuracy improvements. Jin, Williams, Tokar, Waller, et al. (2015) utilise a large set of paired order and point-of-sale data in a retail supply chain to examine the impact of TA on forecast accuracy. They show that it increases forecast accuracy and reduces computational intensity of forecast generation. Luna and Ballini (2011) use TA to predict daily time series of cash withdrawals and find similar or better forecast accuracy to modelling the daily series directly.

Exploring further the impact of TA for demand forecasting Rostami-Tabar et al. (2013) and Rostami-Tabar et al. (2014) derive analytically the optimal aggregation level when the underlying demand process follows AutoRegressive AR(1), Moving Average MA(1), AutoRegressive Moving Average ARMA(1,1) and exponential smoothing is used to produce the forecasts. The choice of forecasting model is motivated by the problem context, where single exponential smoothing is the norm for producing demand forecasts for non-trended and non-seasonal time series. They determine analytically the conditions under which non-overlapping TA outperforms the traditional modelling approach. Using the optimal TA levels, they demonstrate accuracy improvements and show that TA's superiority is a function of the demand process parameters, forecasting method parameters, and aggregation levels. However there are no expressions for more complex ARIMA forms or different forecasting models. This is an important limitation given the prevalence of seasonal and trended demand series in practice. Moreover, it should be noted that the ARIMA type processes can only represent fast moving items. For slow moving items, the consideration of other process such as Integer ARMA (INARMA) processes is relevant (Mohammadipour & Boylan, 2012).

2.2. Multiple temporal aggregation levels

The majority of the aforementioned literature had taken the approach to explore how to best model the time series at a single aggregate level instead of the original that the time series was sampled. Kourentzes et al. (2014) argue that there are two concerns with this approach: (i) for the majority of time series we do not have a way to identify the optimal TA level; and (ii) even if there was one, due to sampling, there is a substantial uncertainty about the underlying process and the appropriate model to apply to a time series. Based on these, they recommend using multiple levels of TA and combining the separate forecasts. This approach not only benefits from managing the modelling risk, but also utilises the established gains of forecast combination (Barrow & Kourentzes, 2016; Blanc & Setzer, 2016). Kourentzes et al. (2014) provide empirical evidence to demonstrate gains over conventional forecasting. Since, modelling with multiple TA levels has been used successfully to intermittent demand, promotional