European Journal of Operational Research 237 (2014) 152–163

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Stochastics and Statistics

'Horses for Courses' in demand forecasting

Fotios Petropoulos [a], Spyros Makridakis [b], Vassilios Assimakopoulos [c], Konstantinos Nikolopoulos [d,*]

[a] Lancaster Centre for Forecasting, Lancaster University, Lancaster, UK
[b] INSEAD Business School, Fontainebleau, France
[c] Forecasting & Strategy Unit, National Technical University of Athens, Greece
[d] Bangor Business School, Prifysgol Bangor University, Bangor, Wales, UK

ARTICLE INFO

Article history:
Received 27 November 2012
Accepted 13 February 2014
Available online 22 February 2014

Keywords:
Forecasting methods
Time series methods
Forecasting accuracy
M-Competitions
Simulation

ABSTRACT

Forecasting as a scientific discipline has progressed a lot in the last 40 years, with Nobel prizes being awarded for seminal work in the field, most notably to Engle, Granger and Kahneman. Despite these advances, even today we are unable to answer a very simple question, the one that is always the first tabled during discussions with practitioners: "what is the best method for *my data*?". In essence, as there are *horses for courses*, there must also be forecasting methods that are more tailored to some types of data, and, therefore, enable practitioners to make informed method selection when facing new data. The current study attempts to shed light on this direction via identifying the main determinants of forecasting accuracy, through simulations and empirical investigations involving 14 popular forecasting methods (and combinations of them), seven time series features (*seasonality*, *trend*, *cycle*, *randomness*, *number of observations*, *inter-demand interval* and *coefficient of variation*) and one strategic decision (the *forecasting horizon*). Our main findings dictate that forecasting accuracy is influenced as follows: (a) for fast-moving data, *cycle* and *randomness* have the biggest (negative) effect and the longer the *forecasting horizon*, the more accuracy decreases; (b) for intermittent data, *inter-demand interval* has bigger (negative) impact than the *coefficient of variation*; and (c) for all types of data, increasing the length of a series has a small positive effect.

1. Introduction

Forecasts are important for all decision-making tasks, from inventory management and scheduling to planning and strategic management. Makridakis and Hibon (2000) advocated: "predictions remain the foundation of all science". To that end, identification of the best forecasting techniques for each data set, or, even, for each series separately, is still the 'holy grail' in the forecasting field, and, as a result, empirical comparisons to this direction are considered very important (Fildes & Makridakis, 1995). Advanced, sophisticated and simpler extrapolation methods could be associated with specific features of data. The development of a protocol for automatic selection of the best tools for resolving each problem, a protocol that would guarantee minimum out-of-sample forecasting error and therefore have a substantial impact on decision making, is the ultimate challenge for researchers and practitioners in the field.

As early as the late 1960s and most of the 1970s, several researchers (Cooper, 1972; Groff, 1973; Kirby, 1966; Krampf, 1972; Levine, 1967; Makridakis & Hibon, 1979; Naylor & Seaks, 1972; Newbold & Granger, 1974) sought to determine the accuracy of various forecasting methods in order to select the most appropriate one(s). In addition, psychologists have been concerned with judgmental predictions and their accuracy, as well as the biases that affect such predictions, for more than half a century (Dawes, 1979; Hogarth, 1987; Kahneman & Tversky, 1973; Meehl, 1954, 1986; Slovic, 1972; Tversky & Kahneman, 1982). Amongst these biases, those affecting forecasting include over-optimism and wishful thinking, recency, availability, anchoring, illusory correlations and the underestimation of uncertainty. In a recent book, Kahneman (2011) describes these and other biases whilst also discussing what can be done to avoid, or minimize their negative consequences and emphatically states: "the research suggests a surprising conclusion: to maximize predictive accuracy, final decisions should be left to formulas, especially in low-validity environments" (Kahneman, 2011, p. 225). Moreover, the growing demand for forecasting big data (e.g. more than 200,000 time series for major retailers) renders the use of automatic statistical procedures necessary.

* Corresponding author. Tel.: +44 (0)1248 38 3796.
E-mail addresses: f.petropoulos@lancaster.ac.uk (F. Petropoulos), spyros.makridakis@insead.edu (S. Makridakis), vassim@fsu.gr (V. Assimakopoulos), k.nikolopoulos@bangor.ac.uk (K. Nikolopoulos).

http://dx.doi.org/10.1016/j.ejor.2014.02.036
0377-2217/

The purpose of this study is to measure the extent to which each of seven time series features (*seasonality*, *trend*, *cycle*, *randomness*, *number of observations*, *inter-demand interval* and *coefficient of variation*) and one strategic decision (the *forecasting horizon*) affect forecasting accuracy. In order to do this, we measure the impact of each of these eight factors[1] by generating a large number of time series – as well as using real data, and measuring the accuracy of the forecasts derived from 14 methods and five combinations of them. Furthermore, a multiple regression analysis is performed to measure the extent to which each of the factors affects the accuracy of each of the time series methods/combinations. The findings of this research could be very useful for practitioners if used for the appropriate selection of the best statistical forecasting practices based on an ex-ante analysis of their data (and their respective features).

This paper is structured as follows: after the literature review (Section 2), the simulation design for fast-moving and intermittent demand data is discussed in Section 3. In Section 4 the accuracy results are presented. Section 5 discusses the findings and Section 6 presents the practical implications for decision makers. Finally, Section 7 concludes and suggests possible avenues for future research.

2. Background literature

Extrapolation models are used very often when facing large amounts of data. Among them, exponential smoothing forecasting approaches were developed in the early 1950s and have become very popular amongst practitioners. Their main advantages are simplicity of implementation, relatively low computational intensiveness and no requirement for lengthy series, whilst being appropriate for short-term forecast horizons over a large number of items. Single Exponential Smoothing (SES – Brown, 1956) uses only one smoothing parameter and is forecasting quite accurately stationary data. Holt's two parameters approach (1957) expands the Single method with a smoothing parameter for the slope, making the method more appropriate for trended data. The Holt-Winters approach (Winters, 1960) is an expansion upon the Holt trended model, which assumes an additive or multiplicative seasonality in the data. Gardner and McKenzie (1985) added a dampening factor ($0 < \varphi < 1$) applied directly on the trend component, resulting in a very successful approach that is often considered the benchmark in many empirical evaluations. Assimakopoulos and Nikolopoulos (2000) proposed the Theta model – a prima facie variation of SES with drift, with the full theoretical underpinnings presented by Thomakos and Nikolopoulos (2014), a method that topped the M3-Competition, the largest empirical forecasting competition to date (Makridakis & Hibon, 2000, Appendix B).

On the other hand, the more complex but quite popular Box–Jenkins methodology (Box & Jenkins, 1970) uses an iterative three-step approach (model identification, parameter estimation and model checking) in order to find the best-fit ARIMA model. To date ARIMA models are still considered the dominant benchmark in empirical forecasting evaluations, and find great popularity among OR researchers in applications spanning from hospitality and production to healthcare and climate forecasting (for e.g. see Broyles, Cochran, & Montgomery, 2010; Cang & Yu, 2014; Cao, Ewing, & Thompson, 2012).

One result that stands for fast-moving data is that combining improves predictive accuracy (Clemen, 1989; Makridakis & Winkler, 1983; Surowiecki, 2005). In addition to this, combining reduces the variance of forecasting errors and therefore the uncertainty in predictions, rendering the selection of combinations less risky than individual methods (Hibon & Evgeniou, 2005). Many recent studies have verified that the combination of methods leads to more accurate forecasts, whilst, at the same time proposing more sophisticated weightings such as the trimmed and Winsorized means (Jose & Winkler, 2008), and the use of information criteria (Kolassa, 2011; Taylor, 2008).

For count data/intermittent data, Croston (1972) proposed decomposing the data into two subseries (demands and intervals) with Syntetos and Boylan (2005) proposing a bias-correction to the Croston's method (Syntetos and Boylan Approximation or SBA). More recently, Teunter, Syntetos, and Babai (2011) suggested a decomposition method that relies on the separate extrapolation of the non-zero demands and the probability to have a demand. This method is very useful in cases of obsolescence. Lastly, simpler approaches, such as Naïve, Moving Averages and SES, have also been quite popular for such data especially among practitioners.

An interesting spin-off from the later intermittent demand literature came from Nikolopoulos, Syntetos, Boylan, Petropoulos, and Assimakopoulos (2011) with the ADIDA non-overlapping temporal aggregation forecasting framework, that although designed and successfully evaluated empirically on count data (Babai, Ali, & Nikolopoulos, 2012), the implications pretty fast span out for fast-moving data as well (Kourentzes, Petropoulos, & Trapero, 2014; Spithourakis, Petropoulos, Babai, Nikolopoulos, & Assimakopoulos, 2011). The proposed framework soon was perceived as a forecasting method "self-improving" mechanism that by changing the data series features through frequency transformation, can help extrapolation methods achieve better accuracy performance. The first theoretical results for the ADIDA framework appeared recently in the literature (Rostami-Tabar, Babai, Syntetos, & Ducq, 2013; Spithourakis, Petropoulos, Nikolopoulos, & Assimakopoulos, in press).

2.1. 'Horses for courses'

Given the plethora of the aforementioned methods, it is now even more unclear: when should each method be used? Many researchers compared the performance of aggregate and individual selection strategies (Fildes, 1989; Shah, 1997; Fildes & Petropoulos, in press). While selecting a single method for an entire data set would make sense for homogeneous data, model selection should be done individually (per series) when we deal with heterogeneous data, as to capture the different features met in each series.

Pegels (1969) presented the first graphical classification for exponential smoothing methods, separating trend from cycle patterns, and also as additive from multiplicative forms. In a simulation study, Adam (1973) evaluated several forecasting models across five different demand patterns, including constant, linear trend, seasonal and step function. His findings indicate that no single model is consistently better than the others, and their performance depends primarily on the demand pattern, the forecasting horizon and the randomness, and secondarily on the selected accuracy metric. Gardner and McKenzie (1988) provided a procedure for model identification in the case of large forecasting applications. Their selected course of action involved the calculation of variances at various levels of differences in data, and using those for classifying the underlying pattern of the time series (constant or trended, seasonal or not seasonal, and so on).

A first attempt for a rule-based selection procedure of the best model derived from Collopy and Armstrong (1992). They proposed a framework that combines forecasting expertise with domain knowledge in order to produce forecasts based on the characteristics of the data. Their procedure consisted of 99 rules and four extrapolation techniques, while 18 time series features were used. A simplified domain knowledge-free version of this rule-based

[1] We use the term 'factor' to refer to both the data features as well as the strategic decision (*forecasting horizon*) which impact on forecasting accuracy will be examined through this study.