Journal of Business Research 69 (2016) 6088–6096

Contents lists available at ScienceDirect

Journal of Business Research

Forecasting intraday call arrivals using the seasonal moving average method

Devon K. Barrow *

Coventry Business School, Coventry University, Priory Street, Coventry CV1 5FB, UK

ARTICLE INFO

Article history:
Received 3 November 2014
Received in revised form 7 December 2015
Accepted 24 June 2016
Available online 25 July 2016

Keywords:
Call center arrivals
Time series forecasting
Seasonal average
Univariate methods
Artificial neural networks
Forecast combination

ABSTRACT

Research into time series forecasting for call center management suggests that a forecast based on the simple Seasonal Moving Average (SMA) method outperforms more sophisticated approaches at long horizons where capacity planning decisions are made. However in the short to medium term where decisions concerning the scheduling of agents are required, the SMA method is usually outperformed. This study is the first systematic evaluation of the SMA method across averages of different lengths using call arrival data sampled at different frequencies from 5 min to 1 h. A hybrid method which combines the strengths of the SMA method and nonlinear data-driven artificial neural networks (ANNs) is proposed to improve short-term accuracy without deteriorating long-term performance. Results of forecasting the intraday call arrivals to banks in the US, UK and Israel indicate that the proposed method outperforms standard benchmarks, and leads to improvements in forecasting accuracy across all horizons.

1. Introduction

Accurate and robust forecasts of inbound calls volumes as a measure of service demand is of primary importance to managing call centers effectively and efficiently, be it for scheduling agents efficiently in 15- or 30-min intervals during the day or within a week, or determining the quantity, and timing of hiring and training (Aksin, Armony, & Mehrotra, 2007; Gans, Koole, & Mandelbaum, 2003). Call centers employ millions of individuals around the world accounting for >70% of all customer-business interactions (Brown et al., 2005). With 60–80% of a call center's operating budget comprising of human resource costs (Aksin et al., 2007) the accurate forecasting of inbound calls, even those corresponding to a single product or service such as a medical emergency hotline, can have substantial socio-economic implications.

Time series forecasting research has recently focused on developing rather sophisticated methods for forecasting inbound call arrivals. However there has been overwhelming evidence (Ibrahim & L'ecuyer, 2013; Tandberg, Easom, & Qualls, 1995; Taylor, 2008a, 2010) that such methods are outperformed by the simple Seasonal Moving Average (SMA) method particularly at longer forecast horizons where capacity planning decisions are made. Despite its attractiveness, the performance of the SMA method has not been systematically evaluated, nor have extensions been investigated. This study evaluates the performance of the SMA method systematically varying the number of seasonal periods included in the average to assess its impact on forecasting accuracy across different data frequencies of 5 min, half-hourly and hourly recorded call arrivals. The SMA method is compared to 'simple' and advanced benchmarks including seasonal ARIMA and the double seasonal Holt-Winters exponential smoothing method of Taylor (2003) forecasting 5 min to two weeks ahead.

A new hybrid forecasting method is proposed which combines the strengths of the simple SMA method, capable of robustly capturing the intraday and intraweek seasonal pattern in intraday call arrivals, and the data driven nonlinear capabilities of ANNs in modelling potential nonlinear and nonparametric features of the residuals (Zhang, Patuwo, & Hu, 1998). Such an approach would allow call center managers the ability to observe both the short- and long-term trends in call arrivals in a single forecast, and facilitate easier use of judgmental adjustments in that it separates out the seasonal weekly and daily fluctuations from the rest of the series highlighting its main components.

Both linear autoregressive (AR) and nonlinear ANNs are evaluated as in practice it is often difficult to determine whether a series is generated from a linear or nonlinear process, and/or whether any one method will produce better forecasts than the other. This is especially true for the case of the three Banks considered in this study, whose service demand are likely affected by both structural and behavioral changes in response to financial and economic stimuli. Data on inbound service demand is obtained from call centers of a US bank (Weinberg, Brown, & Stroud, 2007.), a UK bank (Taylor, 2008a), and a bank in Israeli (Mandelbaum, Sakov, & Zeltyn, 2000). These represent 5 min, half-hourly, and hourly observations of call arrivals respectively and facilitate evaluation of performance

* Corresponding author at: School of Strategy and Leadership, Faculty of Business and Law, Coventry University, Coventry, West Midlands CV1 5FB, UK.
E-mail address: devon.barrow@coventry.ac.uk.

http://dx.doi.org/10.1016/j.jbusres.2016.06.016
0148-2963/

across increasing sampling frequency. These three series have a significant impact on the cost of operations of these call centers, representing a major aspect of inbound call traffic and affecting capacity planning and scheduling decisions. It is hypothesized that by using ANNs, complex autocorrelation structures in the data may be modelled more accurately.

The rest of the paper is organized as follows. In Section 2, a review of the literature on univariate forecasting for intraday arrivals is performed. This is followed in Section 3, by a discussion of the Seasonal Moving Average method and development of the proposed hybrid approach. Section 4 provides a description of the intraday call arrival datasets followed by Section 5 which describes the experimental design and benchmarks method. Section 6 presents the results and findings, while Section 7 discusses briefly the implications of practice. Finally, Section 8 presents a summary and concluding remarks.

2. Univariate methods for forecasting intraday arrivals

The lack of research into time series forecasting methods for call centers first observed by Fildes and Kumar (2002), and detailed by Gans et al. (2003) and Mandelbaum (2006), has led to a recent surge in this area of research. The first empirical evaluation of univariate time series methods for call center arrivals by Taylor (2008a) evaluated several models not previously used for call center forecasting, including the double seasonal Holt-Winters exponential smoothing method and a multiplicative double seasonal ARMA model (Taylor, 2003). These methods were introduced specifically to model the double seasonal pattern inherent in intraday call arrival data[1] (see Fig. 2). Since then, several advanced time series methods have been developed for modelling time series containing such features. These include numerous developments in exponential smoothing (see, for example, Taylor, 2003, 2008b, 2010, 2012; Taylor & Snyder, 2012), ARIMA modelling (see, for example, Antipov & Meade, 2002; Taylor, 2008a), regression including dynamic harmonic (Tych, Pedregal, Young, & Davies, 2002) and discount weighted regression (Taylor, 2010), singular vector decomposition (see, for example, Shen, 2009; Shen & Huang, 2005, 2008a,b), and the use of Gaussian linear mixed-effects models (Aldor-Noiman, Feigin, & Mandelbaum, 2009; Ibrahim & L'ecuyer, 2013).

Despite the focus on more sophisticated methods of forecasting, the findings of Taylor (2008a) suggest that "to use more advanced methods may not be the solution". The study found that for lead times up to about three days ahead, the double seasonal Holt-Winters and the double seasonal ARIMA methods performed well, but beyond short lead times and across all lead times simultaneously, the SMA method with weekly seasonality was best. While SMA with weekly seasonality did not produce the best accuracy in Taylor (2010), primarily because of poor performance at short lead times, it was observed to be the best performing method beyond four days ahead forecasting. Early evidence from Tandberg et al. (1995) in producing forecasts of hourly calls to a regional poison center in New Mexico also found that the SMA method performed well, outperforming Seasonal ARIMA. Further evidence outside of time series methods research was given by Ibrahim and L'Ecuyer (2013) who observed that at relatively long forecasting lead times, the SMA method outperformed a number of statistical models which included, fixed-effects, mixed-effects and bivariate mixed-effects models.

It is therefore surprising that extensions of the Seasonal Moving Average method have not been considered, despite previous findings of residual autocorrelation when fitted to intraday arrivals, a clear indication that further improvements are possible (Brown et al., 2005; Taylor, 2008a). Additionally the method has not been systematically evaluated. This is remarkable given its preferred use in practice over more advanced methods which are difficult to implement, communicate to middle and top management, and which lack transparency. This study assesses the impact of the number of seasonal periods included in calculating the seasonal moving average to better understand the properties of this simple forecasting method. It also proposes a hybrid decomposition approach which in the first step models and forecasts the original series using the SMA method, and in the second step, models and forecasts the residuals of the SMA method using a linear or nonlinear model. The forecasts of the original and residual series are then combined to produce the final forecast. In estimating the nonlinear AR model we consider ANNs as they have shown promise in modelling data containing similar features of intraday and intraweek seasonality (Temraz, Salama, & Chikhani, 1997; Willis & Northcotegreen, 1983). They are flexible not requiring the prespecification of a particular model form and have been successfully employed in numerous forecasting applications (Adya & Collopy, 1998; Hamid & Iqbal, 2004; Zhang et al., 1998). They have however yielded mixed results when modelling intraday call arrivals (see, for example, Taylor & Snyder, 2012; Pacheco, Millan-Ruiz, & Velez, 2009; Millan-Ruiz, Pacheco, Hidalgo, & Velez, 2010), and selecting a single ANN can be difficult owing to the large number of factors which affecting network performance (Zhang & Berardi, 2001). Given the strengths and weaknesses in both approaches, a hybrid approach seems appealing, and may be an effective strategy in practice.

3. Extending the seasonal moving average: A hybrid approach

The most notable paper involving a hybrid approach based on ANNs is by Zhang (2003), combining ARIMA and ANN models, with improved results over both models when used separately. The proposed approach differs in that it is combines the SMA method and ANNs, and is driven by the underlying properties observed in intraday call arrival time series data. It is inspired by research in time series decomposition (Makridakis, Wheelwright, & Hyndman, 2008). In particular decomposition is useful in analyzing underlying latent components of a time series which may have meaningful interpretations (West, 1997) and whose isolation and subsequent independent modelling may enhance forecasting performance by eliminating variability in sub-series. Theodosiou (2011) for example find improvements in forecasting accuracy from the application of the well-known STL decomposition (Cleveland, Cleveland, McRae, & Terpenning, 1990). This is analogous to temporal aggregation and disaggregation which in practice aids the identification of series characteristics across different temporal frequencies as illustrated by Petropoulos and Kourentzes (2014). Hybrid approaches can be similarly used to exploit the benefits of decomposition and combination to improve forecasting accuracy (Timmermann, 2006).

Using this hybrid approach, a time series can be viewed as consisting of both a linear and nonlinear component as follows:

$$y_t = L_t + N_t \tag{1}$$

where L_t denotes the linear component and N_t, the nonlinear component. In the first step, the SMA method is applied to estimate and forecast the linear component containing the intraday and intraweek seasonal patterns. The h-step-ahead forecast using the SMA method is calculated as:

$$y_{t+h} = \frac{1}{k}\sum_{i=1}^{k} y_{t+h-sk} \tag{2}$$

where k is the number of seasonal periods considered in the calculation of the moving average, s is the length of the seasonal cycle and h the forecast horizon. In this study, different values of k are evaluated to determine its impact of forecasting accuracy. The value of s representing either daily or weekly seasonality is chosen to minimize the mean squared error over the training set. For the chosen arrival series, this

[1] Intraday call arrivals exhibit double seasonality and are a subclass of a more general class of time series containing multiple seasonal cycles each of different lengths The term 'cycle' is used to denote any periodically repeating pattern (with variation) in contrast to an economic cycle that has no fixed length (Gould et al., 2008).