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 pre-specification 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).

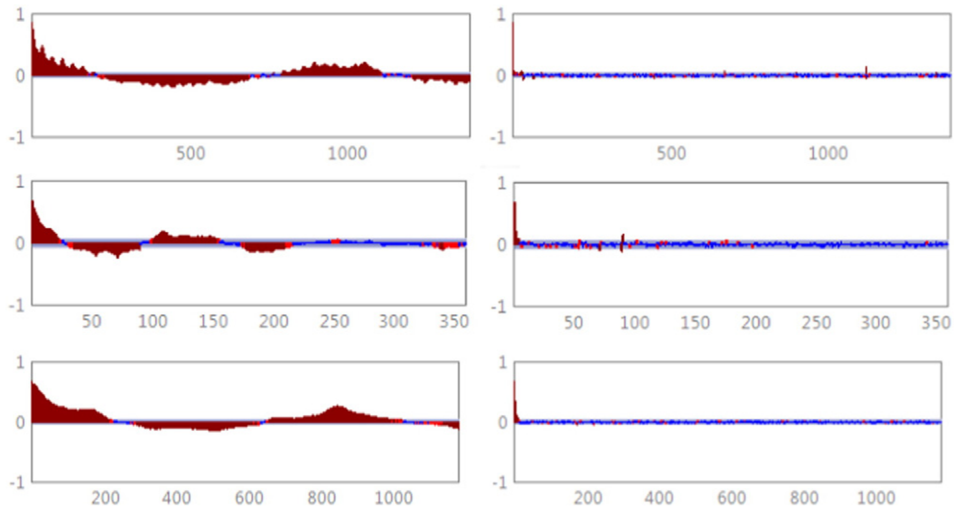

Fig. 1. Autocorrelation [left] and partial-autocorrelation [right] plot of one-step-ahead forecast errors the 5-period seasonal average on the UK bank [top], Israeli bank [middle] and US bank [bottom] datasets.

selection was easily verifiable by observing that both the UK and US series exhibit intraweek seasonal patterns and consequently capturing only daily seasonality was insufficient to model both weekdays and weekends which differ substantially. In contrast for the Israeli series, which excludes weekends, better results were obtained using a daily seasonal cycle. Using this method, the forecast for each lead time is given as the average of call arrivals for the same period of the day or week as the period to be predicted. For example, with 10 weeks (days), the forecasts will be the number of calls arriving for the same period of the week (day) as the period to be predicted, averaged across all previous 10 weeks (days). For hourly data this means that we average the 10 call arrival volumes for the same hourly period corresponding to the previous 10 weeks.

The SMA method is used to produce L_t the one-step ahead forecasted value for time t. In the second step these forecast values are used to calculate one-step-ahead in-sample forecast errors (residuals) of the SMA method. The residual series e_t is given by:

$$e_t = y_t - L_t \quad (3)$$

These residuals generally contain some remaining autocorrelation (Gardner, 1985; Taylor, 2003), evidenced in the residuals of a fitted 5 period SMA shown in Fig. 1. It can be observed that for all three series, the one-step-ahead errors produced tend to go in runs having the same sign, indicating in all cases quite large and positive first-order autocorrelation. This indicates that the forecasts produced by the SMA method are clearly not optimal, and can be further improved. Additionally, all residuals are tested for any nonlinearities using the BDS Test[2] for nonlinearity (Broock, Scheinkman, Dechert, & LeBaron, 1996). The BDS test is used to test for remaining linear dependence and the presence of omitted nonlinear structures in the residuals. For the UK, US and Israeli Bank series, p-value < 0.000 are obtained for residuals of the 5 period SMA method indicating the presence of possible nonlinear structure in the data. A major advantage of the BDS test is that it requires no distributional assumption on the time series data.

[2] The BDS test measures the frequency with which temporal patterns are repeated in a time series counting the number of observations within a specified distance ϵ. The BDS Test statistic measures the closeness of the points with the probability of independent and identical distribution (i.i.d) of the residuals dependent on ϵ and number of past observations. Where the null hypothesis of independent and identical distributions (i.i.d) is rejected, the fitted linear model is deemed misspecified, and provides evidence of nonlinearity. The BDS Test is implemented using the R Software and the fNonlinear package with ϵ set to 0.5, 1, 1.5 and 2 standard deviations of the data set and embedding dimensions 1 to 5.

To correct for the autocorrelation observed in the data, and to investigate whether there are any benefits of applying a nonlinear approach over a linear approach in step 2, both a linear autoregressive model of order p, AR(p), and a Multilayer Perceptron (MLP) neural network of the form NAR(p) are evaluated in modelling the residuals. Details of both models and their setup are provided in Appendix A. In either case, the model for the residuals is of the form:

$$e_t = f(e_{t-1}, e_{t-2}, \ldots, e_{t-p}) + \varepsilon_t \quad (4)$$

where f is a linear or nonlinear function, and ε_t is the random error. Finally, the decomposed forecasts of the SMA method and that of the residuals are summed to obtain the final combined forecast as follows:

$$y_t = L_t + N_t \quad (5)$$

where N_t is the forecast of the possibly nonlinear SMA residual component. While it may be more efficient to estimate all parameters for this method in a single stage (Chatfield, 1985), a two-stage estimation approach is adopted to maintain simplicity, in terms of calculation and optimization of model parameters, and transparency of the method to enhance decision making and the use of judgment.

4. Call center arrivals data

Three time series of intraday call arrivals known to exhibit complex seasonal patterns are considered. Such series tend to have a daily seasonal pattern or intraday cycle, a weekly seasonal pattern or intraweek cycle, and where multiple years of observations are available, an annual seasonal pattern. A plot of all three series is shown in Fig. 2. The first (from the bottom up) consists of hourly data corresponding to regular calls from 1 August to 25 December inclusive, taken from a small call center at one of Israel's banks (Mandelbaum et al., 2000). The call center operates 18 h per day from 6 A.M. to 12 P.M. and is open 5 days per week (see Table 1). Fig. 1[bottom] presents the final four weeks of the series, which shows no apparent trend, and illustrates an intraday seasonal cycle containing $s_1 = 18$ periods, and a possible intraweek seasonal cycle of $s_2 = 5 \times 18 = 90$ periods excluding Saturdays, and including holidays. The first 14 weeks of the series are used for method estimation, while the remaining seven weeks are used for out-of-sample forecast evaluation.

The second series consists of half-hourly arrivals at the call center of a major retail bank in the United Kingdom (Taylor, 2008a). This call center operates 7 days per week and is open 16 h (32 half-hours) per day

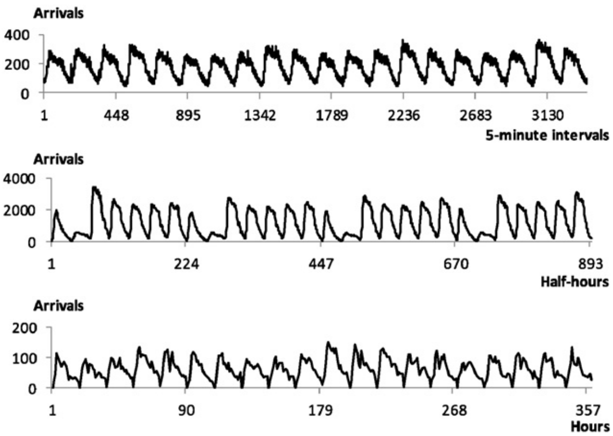

Fig. 2. Intraday call arrival time series for the US bank [top], UK bank [middle] and Israeli bank [bottom].

(see Table 1). The final four weeks of the series is shown in Fig. 2 [middle].The series exhibits no apparent trend, but contains both an intraday seasonal cycle, $s_1 = 32$, and also a repeating intraweek seasonal cycle, $s_2 = 7 \times 32 = 224$ periods. The entire time series of 36 weeks, from 3 January 2004 to 10 September 2004 both days included is considered. The first 24 weeks of data were used to estimate various model parameters, while the remaining 12 weeks provide the holdout for the post-sample forecast evaluation.

The final dataset comes from a large North American commercial bank, and consists of 10,140 observations of 5-min interval call arrivals into a retail bank for the first 12 weeks (excluding weekends) starting from 3 March 2003 to 28 May 2003 inclusive (De Livera, Hyndman, & Snyder, 2011; Weinberg et al., 2007). The call center opens 5 days per week and operates from 7 A.M. to 9.05 P.M approximately 14 h and 5 min (or 169 5-min intervals). Fig. 2 [Top] shows the final four weeks of the call arrival series which contains both an intraday seasonal cycle $s_1 = 169$ periods, and an intraweek seasonal cycle of $s_2 = 5 \times 169 = 845$ periods. The first 7605 observations or 9 weeks are used to estimate model parameters, and the remaining 2535 observations or 3 weeks, for post-sample forecast evaluation. Table 1 summarizes the number of 5-min intervals, half-hours and hours in the estimation and evaluation sample for all series. It should be noted that the size of the in- and out-of-sample datasets were selected to remain consistent with prior studies. In the case of the UK and Israeli Bank series a setup similar to that of Taylor (2008a) is used, while for the US Bank series the setup of De Livera et al. (2011) is adopted.

5. Preprocessing, parameter estimation and evaluation

5.1. Data preprocessing

All time series are modelled without prior smoothing of bank holidays or other “special days”. This approach was deemed reasonable as in practice, univariate methods are often required to produce robust baseline forecasts in the presence of such days. It is then expected that experts, if required, will overlay their experience and knowledge of the demand series in accounting for the impact of such special days

Table 1
Summary and description of the call arrivals for the Israeli and UK Bank Call Center.

	Days opened per week	Opening hours	Recorded interval	Size of estimation sample	Size of evaluation sample
Israeli Bank	5	6 A.M. − 12 P.M.	Hourly	1260	630
UK Bank	7	7 A.M. − 11 P.M.	Half-hourly	5376	2688
US Bank	5	7 A.M. −9:05 P.M.	5-min	7605	2535

such as bank holidays. Other studies in modelling arrival series have applied a logarithmic transformation prior to model estimation in order to reduce the impact of heteroscedasticity (Brown et al., 2005; Taylor, 2008a). The impact of this transformation is assessed for both the UK and US series. If it resulted in the lowest in-sample mean squared error the transformation was applied, otherwise the series was modelled in its original form. For the Israeli Bank data, it was not possible to apply the log transformation due to periods with no (zero) call arrivals. Instead for this series a square root transformation used in the study by Taylor (2008a) and Brown et al. (2005) was used to reduce heteroscedasticity.

5.2. Model estimation

5.2.1. The seasonal moving average

The length of the seasonal moving average is assessed to determine its impact on forecasting accuracy. For the UK and Israeli bank series k the length of the moving average is set to 5, 10 and 15 periods with seasonality s equal to 224 and 18 to model weekly and daily seasonality, respectively. Finally, for the US Bank data averages of lengths 2 and 5 are considered due to the limited data available. For this series seasonality s is set to 845 representing weekly seasonality. For the UK and US series, the choice of seasonal cycle was consistent with the properties of the arrivals series, both of which exhibited evidence of an intraweek seasonal cycle. For the Israeli series, which excluded weekends, better results were obtained using a daily seasonal cycle.

5.2.2. The multilayer perceptron

The most commonly applied artificial neural network, the Multilayer Perceptron (MLP), a feedforward artificial neural network is used in the study. A single MLP architecture is used to forecast all time series, using two hidden nodes and a single output node with an identity function producing multistep forecasts recursively through one-step-ahead forecasts. In selecting the network inputs for modelling the residuals of the SMA method in the hybrid approach, a stepwise regression is used having maximum order of 5 considering the maximum order of the AR term in the seasonal ARIMA. In modelling the original series using only the MLP, a mixed approach is adopted based on stepwise selection which has proven a suitable contender for high-frequency time series (Crone & Kourentzes, 2010). For each series, the input lags with the highest statistical significance are identified using partial autocorrelation analysis. This reduced subset of selected lags is then used as input to a stepwise regression to select the final set of inputs for the network, in effect pre-filtering the search space. Each time series is modelled directly without prior differencing. Inputs were linearly scaled into the interval of $[-0.5, 0.5]$ to allow headroom for possible non-stationarity prior to training. The training algorithm used is the standard backpropagation algorithm, minimizing the mean square error up to a maximum of 1000 epochs. The algorithm requires setting a learning rate $\eta = 0.02$ and momentum parameter $\mu = 0.7$. As neural network training performed in this manner is subject to the local minima problem, where the nonlinear optimisation gets trapped in the local minimum of the error surface potentially resulting in poor quality results, training is initialised several times with different random starting weights and biases to explore the error surface more fully. The best training initialisation is retained as the final model having the lowest in-sample mean squared error. While neural network model averaging is generally advocated in preference to model selection (Kourentzes, Barrow, & Crone, 2014), there was no clear distinction between the results of model averaging and model selection. The good performance of model selection on the select call arrival series has been attributed to the large sample size available for training such series (due to the high frequency nature of the data) and the reduced degrees of freedom (from fewer required lagged inputs) as a result of prior removal of daily and weekly seasonal effects – an advantage of the proposed Hybrid method. Additionally due to its relative simplicity we report results based on model selection.

5.2.3. Multiplicative seasonal ARIMA

In addition to the methods previously described three benchmarks are evaluated. The first is the multiplicative seasonal ARIMA which has appeared in many studies, and is particularly good at short-term forecasting. Bianchi, Jarrett, and Hanumara (1998) for example found in their study that ARIMA modelling outperformed both additive and multiplicative versions of Holt-Winters exponential smoothing. The model is often written in short form as $ARIMA(p,d,q)\times(P,D,Q)_s$ where p, q and P, Q, are the orders of the autoregressive and moving average terms of the non-seasonal and seasonal components respectively, and d, D are the orders of differencing (Dalrymple, 1978). An ARIMA model is fitted for all three series using the estimation sample in Table 1. In estimating the seasonal ARIMA model, the seasonal length producing the lowest in-sample forecast error according to mean absolute error (MAE) is selected. This simple selection easily discriminates between daily and weekly seasonality given the large difference in performance between the two seasonal cycles. Parameters of the model were estimated using maximum likelihood based on the standard assumption of Gaussian distributed errors. The final model was selected using the Akaike information criteria (AIC). The orders of the ARIMA model selected for each series are given in Table 2.

5.2.4. Holt-Winters exponential smoothing

Forecasts were produced using two Holt-Winters methods, the standard Holt-Winters for multiplicative seasonality, and the double seasonal Holt-Winters method or in short, Double Holt-Winters. In order to estimate the smoothing parameters α (level), γ (trend), δ (seasonal period s_1), ω (seasonal period s_2) and ϕ (AR adjustment) for all three series, the estimating the procedure of Taylor (2003) is used which minimizes the sum of squared errors on the estimation sample, in a single procedure. This is achieved through the implementation provided in the forecast package for R (Hyndman, 2010).

For the standard Holt-Winters method capable of modelling only single seasonality the same procedure previously described is used selecting for each time series, the seasonal length, s, representing daily or weekly seasonality, as the one with the lowest in-sample MAE. The final model chosen was then either the Holt-Winters for daily seasonality or Holt-Winters weekly seasonality. For initializing the smoothed parameters for the level, trend and seasonal components in the standard Holt-Winters method, the simple average of the first two weeks of observations is used as a heuristic (Hyndman, Koehler, Snyder, & Grose, 2002). This is to reduce the impact of known over-parameterization issues when optimizing exponential smoothing for high-frequency double seasonal call arrival series including the potentially large optimization problem from the increased number of initial seasonal values to be estimated (De Livera et al., 2011). The estimated parameters for each of the three call arrivals series for the standard, and double seasonal version of Holt-Winters are shown in Table 3.

5.3. Forecast evaluation

The mean absolute error (MAE) and the symmetric mean absolute percentage error (SMAPE) are used as measures of out-of-sample forecast accuracy. The MAE is used as it allows the direct comparison of forecasting accuracy and the estimation of improvements in accuracy while SMAPE is a scale independent measure which facilitates the reporting of average performance across time series. SMAPE is also selected to facilitate comparison with prior studies such as Taylor (2008a) who use MAPE. SMAPE is preferred to MAPE being more symmetric in that it gives more equal weighting to positive and negative errors (Armstrong & Collopy, 1992). For the UK, US and Israeli Bank series, the holdout out-of-sample evaluation period is set to 12 weeks, 3 weeks and 7 weeks respectively, being the most recent observations. For the UK Bank series, the forecast lead time is set to one half-hour ahead up to 14 days ahead (or 448 half-hours). A rolling origin forecast evaluation is performed (without model re-estimation) producing trace forecasts for each lead time, from each observation in the out-of-sample period. For the UK Bank series this yields 2240 half-hour time origins for a total of 1,003,520 forecasts. For the US Bank series, the average error is calculated across lead times from 5 min to one day ahead yielding 1690 multiple-step-ahead out of sample predictions across multiple time origins and a total of 1,428,050 forecasts. Finally for the Israeli Bank series forecasts are produced from 1 h ahead up to 2 weeks ahead generating forecasts from 450 hourly time origins to create 81,000 forecasts. This yields a large set of forecasts, and forecast error measurements for each method, and for each horizon, and hence a more reliable and robust estimation of the empirical distribution of errors for different horizons (Tashman, 2000). In addition, the Giacomini and White Conditional (GW) test with the null hypothesis of equal forecast accuracy is used to compare the forecast accuracy of competing methods in a multiple pairwise comparison (Giacomini & White, 2006) of the select best SMA, hybrid and benchmark methods. For each time series and pair of methods the out-of-sample forecast errors for the relevant h-step-ahead forecasts is compared to assess performance at the longest horizon. The Giacomini and White Conditional Test directly accounts for the effects of estimation uncertainty on forecast performance in contrast to unconditional tests such as the Diebold Mariano Test which do not take into account differing model complexities (Giacomini & White, 2006). It is also chosen over the Diebold Mariano Test as it allows a unified treatment of both nested (e.g. SMA methods of different lengths) and non-nested models (SARIMA and MLP). Results of the GW test together with the large number of out-of-sample errors generated using rolling origin forecasting are deemed sufficient to ensure valid and reliable results (albeit only for the assessed datasets).

Table 2
Orders of the fitted $ARIMA(p,d,q)\times(P,D,Q)_s$ models for each of the three call arrivals series.

	p	d	q	P	D	Q	s
Israeli Bank	2	0	0	2	1	2	18
UK Bank	2	1	1	2	1	0	224
US Bank	1	0	2	0	1	0	845

Table 3
Smoothing parameters of the fitted exponential smoothing methods for each of the three call arrivals series.

	α	γ	δ	ω	ϕ
Standard Holt-Winters					
Israeli Bank ($s = 18$)	0.266	0.272	–	–	–
UK Bank ($s = 224$)	0.000	0.166	–	–	–
US Bank ($s = 845$)	0.201	0.101	–	–	–
Double Holt-Winters					
Israeli Bank ($s = 18$)	0.032	0.002	0.023	0.304	0.383
UK Bank($s = 224$)	0.023	0.000	0.074	0.291	0.680
US Bank($s = 845$)	0.119	0.000	0.046	0.201	0.277

6. Experimental results

6.1. Overall forecasting accuracy

Table 4 summarizes the SMAPE and MAE for the UK, Israeli and US Bank series presented as averages across lengths of averages for the SMA and hybrid forecasting methods, SMA with MLP (SMA_{MLP}), and SMA with AR (SMA_{AR}). Each column summarizes the lead times for a given series in terms of short, medium and long lead horizons. For the UK series the short horizon represents 1–5 days ahead, medium horizon 6–10 days ahead, and long horizon 11–14 days ahead. Each column is

Table 4
Mean SMAPE and MAE for the UK, Israeli and US bank series.

Forecast horizon	UK Bank				Israeli Bank				US Bank			
	Short	Medium	Long	All	Short	Medium	Long	All	Short	Medium	Long	All
SMAPE												
SMA	**9.34%**	**9.47%**	**9.31%**	**9.38%**	**23.92%**	**23.70%**	**24.23%**	**23.95%**	9.01%	8.99%	8.93%	8.98%
SMA_{MLP}	9.65%	9.97%	9.68%	9.77%	24.62%	23.91%	24.63%	24.41%	**8.28%**	**8.64%**	**8.89%**	**8.64%**
SMA_{AR}	9.35%	9.52%	9.62%	9.49%	27.44%	28.16%	28.72%	28.04%	9.24%	9.22%	9.17%	9.21%
MAE												
SMA	102.18	103.89	**100.82**	102.40	11.73	**11.55**	**11.91**	**11.73**	15.36	15.22	15.11	15.21
SMA_{MLP}	101.40	105.64	101.65	102.99	**11.64**	11.62	12.01	11.74	**13.82**	**14.69**	**15.10**	**14.65**
SMA_{AR}	**98.59**	**103.04**	104.23	**101.79**	13.01	13.30	13.56	13.26	15.58	15.47	15.38	15.46

Note: Forecast errors in boldface indicate the best performing method at each time horizon.

the average of half-hourly forecast errors corresponding to the respective horizon. For example, the column heading "Short" contains the average of the MAE for lead times of 1 half-hour to 5 days ahead (i.e. 160 half-hour periods ahead). The final column provides the average MAE across all lead times, that is, 14 days ahead. For the Israeli series short horizon represents 1–4 days ahead, medium horizon 5–7 days ahead, and long horizon 8–10 days ahead, each column being the average of hourly forecast errors corresponding to the respective horizon. The column heading "Short" therefore contains the average of the MAE for lead times of 1 h to 4 days ahead or equivalently 1–72 h ahead. The final column provides the average MAE across all lead times up to 10 days ahead. Finally for the US series the column heading "Short" contains the average of the MAE for lead times of 5 min to 2 h 35 min ahead indicating short horizon, medium representing 2 h 40 min to 10 h ahead, and long horizon representing 10 h 5 min ahead to 14 h ahead. The final column provides the average MAE up to 1 day ahead. Values in bold indicate the best-performing method for each horizon averaged across seasonal lengths 2, 5, 10 and 15.

Results indicate that on average the SMA method outperforms the hybrid approach on the Israeli Bank series while for the US Bank series, the hybrid SMA_{MLP} has best accuracy. For the UK Bank series results are somewhat inconsistent with SMAPE ranking SMA as best while MAE suggests that the hybrid SMA_{AR} method provides best results. These results suggest that for the UK and Israeli series, the SMA method is somewhat robust to the length of the seasonal moving average, while for the US Bank series where lengths 2 and 5 are considered, using the MLP outperforms both methods particularly at short horizons. These results while providing a good summary, are somewhat inconsistent and suggest the need to drill further to consider the impact of the moving average length on accuracy of both approaches.

6.2. Impact of the length of the seasonal average

The impact of the length of the seasonal moving average is evaluated for both the SMA method and the proposed hybrid methods, and results presented in Table 5 using the MAE to directly measure the accuracy on each time series. Values in bold highlight for each method the best choice of moving average length. For example, the best accuracy using SMA at short horizons on the UK Bank series is achieved using a seasonal moving average of length 5 giving an MAE of 100.81. Values marked in bold and underline highlight for each series and forecast horizon, the best performing method overall. The method having the lowest forecast error on the UK series at short horizons is therefore the hybrid SMA_{AR} of length 5 (MAE = 97.56).

Results shown in the final columns of Table 5 labeled "All" show that the best method across all lead times together is the hybrid SMA with MLP Adjustment which for each series records the best accuracy (UK MAE = 100.80; Israeli MAE = 11.47; US MAE = 14.50). The MLP adjustment provides across all lead times, an improvement over using the SMA only forecast and is always more accurate than the hybrid SMA_{AR}. This is subject to the right choice of k, the length of the average. Results on the UK and Israeli Bank series suggest that longer averages tend to perform better at longer horizons with the length 15 average ranking best at long horizons for all methods except the SMA_{MLP}. Similarly at short to medium horizons, shorter averages containing more recent information tend to have the best accuracy. This makes intuitive sense as at shorter lead-times a shorter average will give greater importance to more recent and up to date changes in inbound call demand, while the further out the forecast, the more long term historic trends in demand become important.

Results also suggest that the best length SMA in terms of accuracy does not always produce the hybrid forecast having the best accuracy.

Table 5
Mean MAE for the UK, Israeli and US bank series.

Forecast horizon	UK Bank				Israeli Bank				US Bank			
	Short	Medium	Long	All	Short	Medium	Long	All	Short	Medium	Long	All
SMA												
Length 2	–	–	–	–	–	–	–	–	**15.18**	**15.15**	15.17	**15.16**
Length 5	**100.81**	**103.47**	101.49	**101.95**	12.49	11.97	12.71	12.40	15.54	15.28	**15.05**	15.26
Length 10	102.84	104.35	100.89	102.82	**11.19**	**11.34**	11.77	11.41	–	–	–	–
Length 15	102.88	103.86	**100.07**	102.43	11.53	11.36	**11.26**	**11.40**	–	–	–	–
SMA_{MLP}												
Length 2	–	–	–	–	–	–	–	–	14.33	14.82	**15.04**	14.79
Length 5	**98.00**	**103.28**	**101.18**	**100.80**	12.41	12.04	12.82	12.42	**13.31**	**14.56**	15.17	**14.50**
Length 10	102.34	106.39	101.57	103.56	**11.14**	11.41	11.87	11.44	–	–	–	–
Length 15	103.86	107.24	102.21	104.60	11.37	**11.40**	**11.34**	**11.37**	–	–	–	–
SMA_{AR}												
Length 2	–	–	–	–	–	–	–	–	**15.43**	**15.31**	**15.26**	**15.32**
Length 5	**97.56**	**102.95**	104.92	**101.59**	13.85	13.74	14.31	13.96	15.73	15.63	15.50	15.61
Length 10	99.04	103.18	104.10	101.96	12.97	13.57	13.81	13.40	–	–	–	–
Length 15	99.18	103.00	**103.68**	101.83	**12.21**	**12.59**	**12.55**	**12.42**	–	–	–	–

Note: For each SMA method, the seasonal length with the lowest forecast error at each forecast horizon is highlighted in boldface. The method with the lowest forecast error overall at each forecast horizon is underlined.

For example on the US Bank series, the SMA length 5 has best accuracy at long horizons (MAE = 15.05), while at the same horizon the choice of moving average of length 2 produces best results for the SMA_{MLP} (MAE = 15.04) and SMA_{AR} (MAE = 15.26) methods. When considering average performance across all horizons, rankings of the SMA and hybrid methods are however more consistent suggesting that the best choice of SMA will lead to the best performing hybrid approach overall.

While in-sample results are not reported here, it is also observed that in- and out-of-sample errors of the SMA and hybrid methods are rather consistent, meaning that the length of the SMA with the lowest error on the in-sample training data also on average has the lowest out-of-sample forecast error. This finding is useful for method selection, as improvements in forecasting accuracy from the appropriate selection of the length of the moving average ranges from 1% to 4% for the UK Bank series, 4% to 8% for the Israeli Bank series, and 0% to 2% for the US Bank series. At shorter horizons the gains in the hybrid method over the SMA is even greater. This is further illustrated in the next section which compares the performance of the best SMA and hybrid methods.

6.3. Seasonal moving average versus the hybrid method

Having evaluated the impact of the length of the seasonal moving average, the best performing SMA and hybrid methods are compared for each time series. The best SMA and hybrid method is selected in-sample minimizing the MAE. Results are shown in Fig. 3 for the SMA, SMA_{AR} and SMA_{MLP} methods for all three banks. For the UK Bank series the length of moving average with best in- and out-of-sample performance is 5, and the methods are denoted $SMA(5)_{MLP}$ and $SMA(5)_{AR}$ for the MLP and AR hybrid forecasts, respectively. The hybrid approach with MLP and AR adjustment are especially effective at improving the short-term forecasting accuracy of the SMA method (see Fig. 3 [left]), without diminishing long-term accuracy. Forecasting one day ahead, a 5% reduction is noted over the SMA method when using SMA_{MLP} while across all horizons, SMA_{MLP} and SMA_{AR} are never worse than the SMA method.

Fig. 3 [middle] which shows for the Israeli Bank series the improvements over the SMA method from implementing the AR and MLP adjustments, show that the SMA_{MLP} forecast is always better, or just as good as the original SMA forecast, while the AR adjusted SMA_{AR} forecast is significantly worse beyond horizons of 2 h ahead. This degradation in performance in possibly due to the residuals of the Israeli series which relative to the US and UK Bank series are less well behaved (see Fig. 1). The AR model suffers more as it is unable to respond to these unexplained variations in the residual series resulting in a poorly estimated model. In contrast the MLP model given its flexibility for modelling such structures, does not suffer from similar degradation in performance.

For the case of the US Bank data it can be observed that the MLP adjustment provides a significant improvement over the SMA method at lead times of 5 min to approximately 12 h ahead while at lead times of 12 to 14 h the SMA method is best. This results in an overall reduction in MAE of 5% from using the hybrid MLP. In contrast, the AR adjustment in SMA_{AR} results in an increase in forecast error a lead times beyond approximately 5 h ahead. Scatterplots of the lagged residuals reveal no noticeable nonlinearity, however given similar inputs, the MLP performance may be explained by its powerful adaptive learning, further enhanced by prior smoothing of the SMA method reducing overfitting to any noise in the data. The performance of SMA_{AR} relative to SMA suggests possible sub-optimality in separately estimating the seasonal moving average component and application of a linear autoregressive model. This issue is explored in the next section.

6.4. Benchmark comparisons

In this section, the performance of the best SMA and hybrid methods are compared to those of standard benchmarks for intraday call arrival series. Results are shown in Fig. 4. For the case of the US series (see Fig. 4[right]), observe that up to an hour ahead, HWT exponential smoothing and MLP are best. However beyond that, the 5 period SMA method with an MLP adjustment is best (MAE = 14.50). The method performs well at both short and long lead times indicating benefits of combining both the SMA method and MLP. The Double Holt-Winters (Double HWT) fitted based on the R Forecast Package is observed to perform particularly poorly at lead times beyond 1 h ahead. The model has been observed in empirical applications to suffer from optimization problems due to the large number of initial seasonal values to be estimated when the seasonal cycle is large as is the case with the 5-min US Bank data (De Livera et al., 2011).

The MAE achieved on the UK series (see Fig. 4[left]) are substantially larger than those of the US bank series which receive a much smaller volume of calls. The results are similar to those obtained on the US Bank series, with the SMA_{MLP} method of moving average length 5 outperforming all methods across all lead times together. The SMA method performs well in comparison to all methods (length 5, MAE = 101.95) and is only outperformed by the hybrid approach with MLP (length 5, MAE = 100.80) and by the HWT method (MAE = 101.10).

When compared to the SMA method and its proposed extensions, SMA_{MLP} and SMA_{AR}, both the single MLP and seasonal ARIMA methods perform rather poorly across all horizons from short to long. At short horizons the Double HWT method (MAE = 106.05) performs better than the MLP and seasonal ARIMA, however beyond four days its performance degrades quickly and is outperformed by the MLP. It is also noted that for the UK series, the fitted HWT model approximates a seasonal moving average having alpha parameter value of 0.000 and gamma parameter value of 0.166 as per Table 3. Consequently the behavior of HWT is similar to that of the SMA method which for this series has rather robust performance across lead times suggesting the presence of a strong deterministic seasonal component. Similar results were obtained by Taylor (2008a,b) for the case of the SMA method.

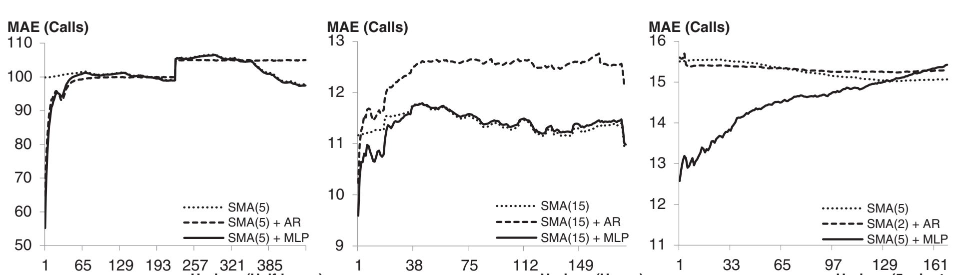

Fig. 3. Mean MAE for the UK (left), Israeli ([middle] and US (right) bank series.
Note: The pairwise differences in forecast errors at the longest horizon for all methods and for all series are found to be statistically significant at the 0.05 level of significance using the Giacomini and White Conditional (GW) test of predictability (Giacomini & White, 2006).

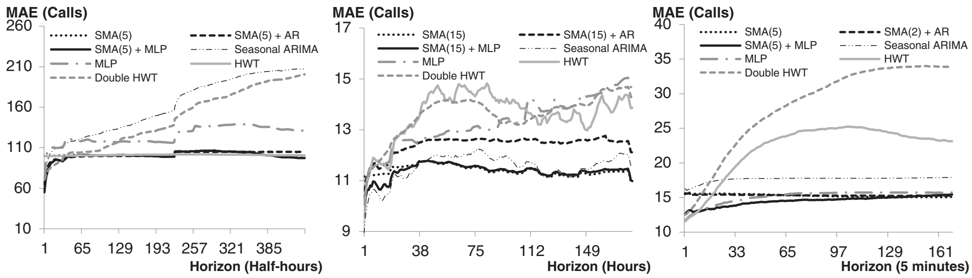

Fig. 4. Mean MAE for the UK (left), Israeli ([middle) and US (right) Bank series.
Note: The pairwise differences in forecast errors at the longest horizon for all methods and for all series are found to be statistically significant at the 0.05 level of significance using the Giacomini and White Conditional (GW) test of predictability (Giacomini & White, 2006).

Finally, for the Israeli Bank series it is observed that the best performing methods across all lead times considered together is the $SMA(15)_{MLP}$ hybrid method, followed by the SMA method, and seasonal ARIMA. The findings for this series are similar in ranking to those of Taylor (2008a), with the SMA and seasonal ARIMA methods ranking similarly albeit with higher errors possibly due to not smoothing out special days and/or the higher frequency (hourly versus half-hourly) at which the time series is modelled in this study. Although beyond the scope of this study, this raises interesting questions about the impact of time series frequency on modelling and forecasting accuracy, and the impact of temporal aggregation and disaggregation. When specific lead times are considered, it can be observed that from 1 h to two days ahead, the seasonal ARIMA model is slightly better than the SMA method, however beyond that, the $SMA(15)_{MLP}$ method ($MAE = 11.37$) outperforms the seasonal ARIMA ($MAE = 11.53$), highlighting the strength of this approach at long-term forecasting (see figure [left]). Results on all time series suggest that modelling intraday call arrivals using the hybrid approach with an MLP network performs well.

7. Implications and practical considerations

Research and practice suggest that the Seasonal Moving Average method performs well relative to more sophisticated methods of forecasting for intraday call arrivals. The findings of this work consolidate existing evidence, and provide new evidence that the SMA method performs particularly well at forecasting intraday call arrivals with the added advantage of being easy to implement. It is observed that the length of the seasonal moving average chosen is an important determinant of performance of the SMA method. At long horizons where long term trends are important, it is recommended to use a longer seasonal moving average shown to outperform the shorter average. Similarly at short horizons a shorter moving average reflecting more recent changes is recommended. This is an important finding for company's currently using this method of forecasting and should assist in identifying the optimal window size of the SMA method.

Results show that in- and out-of-sample results of the Seasonal Moving Average method and related hybrid methods are consistent, suggesting that selection of the length of the moving average should be robust. One approach for doing this is evaluated in this study. The length of the SMA method is determined by minimizing the in-sample MAE. For all time series this simple selection approach consistently obtained best performance in- and out-of-sample.

Using this simple selection method, the hybrid method combining the strengths of the SMA method and the MLP network is able to consistently improve upon the SMA at short and long horizons. The proposed approach using MLP is found to be more robust than the alternative which fits a linear AR model. Where the arrival data is not well behaved and the residuals complex, fitting an MLP neural network yields better performance than the linear AR model. Using the hybrid approach call center managers obtain a single forecast which does well both in the short and long term, and which on average outperforms the SMA and benchmarks for the dataset considered. Identifying the short and long term seasonal movements of interest to management also helps to enhance judgmental adjustments by distinguishing the main sources of variance in the time series, and providing a more informed, clear and robust baseline forecasts on which to base judgment (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009).

8. Summary and concluding remarks

The forecasting of intraday call arrivals exhibiting multiple seasonality requires forecasts which perform well both in the short-term in order to schedule agents effectively, and also in the long-term in order for capacity planning. Recent developments in forecasting this type of data, have led to the development of several advanced methods capable of accounting for intraday and intraweek seasonal patterns. No single time series method has however emerged that is best across all time series. While seasonal ARIMA and Double Holt-Winters have been shown to perform well in the short-term, in the long-term their performance rapidly deteriorates.

This paper provides a systematic evaluation of the Seasonal Moving Average method considering three real time series of inbound demand to call center of major banks in the UK, Israel and the US. It provides insights on selecting the best length for the average and provides an easy to implement way of doing so. A hybrid method is introduced which first models the original arrivals series using the Seasonal Moving Average method, and combines the resulting forecast with the forecast of the residuals using either a linear autoregressive model, or a nonlinear MLP neural network. The results show that for the UK and US arrival series, the proposed method outperformed both the seasonal ARIMA and Double Holt-Winters methods, across all lead times. For the Israeli Bank series it outperforms the seasonal ARIMA in the medium to long term. In nearly all cases the MLP adjustments leads to improvements over the Seasonal Moving Average method. The gains in forecasting accuracy turn out to be substantial, in particular for the US and UK series, which have substantially more observations and are higher volume. Comparing the approaches across these three time series also ensured a robust evaluation across increasing frequencies, from 5 min, to half-hourly and hour observations, producing rather consistent results in the favor of the hybrid method and the Seasonal Moving Average method.

From a practical perspective, the decomposition of the forecast into a simple seasonal moving average which constitutes most of the variance in the series, adds to the ease of implementation with the most complex requirement being the neural network model for which we restricted our study to standard settings. In addition, for staffing purposes, improvements in forecasting accuracy at short and long lead times and the reduced variance in the average forecast error across lead times compared to other methods as observed in Fig. 4, should improve the robustness of staff schedules and deliver substantial cost savings. This would suggest an evaluation of these methods beyond forecasting accuracy as future research. Alternatively this work, together with the findings of Taylor (2008a) would suggest that there are likely benefits from exploiting the advantages of simple and advanced approaches through forecast combination and hybrid methodologies based on time series decomposition.

Appendix A. Linear and nonlinear AR model setup

Further details on the setup and definition of the linear AR and nonlinear AR models are provided in this section. The linear AR(p) model used to fit the residual series takes the form:

$$e_t = \sum_{j=1}^{p} \lambda_j e_{t-j} + \epsilon_t \tag{A.1}$$

where e_t is an estimated of the residuals e_t defined in Eq. 2 being the 1-step-ahead in-sample errors of the seasonal average model, λ_j is the AR coefficient of the j^{th} lag, and $\epsilon_t \sim N(0, \upsilon)$. An MLP is employed to estimate nonlinear AR models of the residuals and to produce the valuee_t, an estimate of the one-step ahead forecast errors as follows:

$$e_t = f(X, w) = \beta_0 + \sum_{k=1}^{H} \beta_h g\left(\gamma_{0i} + \sum_{i=0}^{I} \gamma_{hi} p_i\right) + \varepsilon_t \tag{A.2}$$

with t denoting the point in time and p_i the inputs which are time lagged observations of the residual series e_t. The network parameters are denoted as weights $w = (\beta, \gamma)$ with $\beta = [\beta_1, \ldots, \beta_H]$ and $\gamma = \gamma_{11}, \ldots, \gamma_{HI}$ corresponding to the output and hidden layer respectively, and β_0 and γ_{0i} the biases of each neuron. Parameters $I = (1, \ldots, I_{\max})$ and $H = (1, \ldots, H_{\max})$ specify the number of input and hidden nodes of the network architecture with $I_{\max}$ and $H_{\max}$ the maximum number of input and hidden nodes respectively, while $g(\cdot)$ is a non-linear transfer function in the hidden layer nodes, conventionally set as either the sigmoid logistic or the hyperbolic tangent function (Zhang et al., 1998). The time series is modelled by adjusting network parameters to minimize the mean squared error on the training data.

References

Adya, M., & Collopy, F. (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, *17*, 481–495.

Aksin, Z. N., Armony, M., & Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, *16*, 665–688.

Aldor-Noiman, S., Feigin, P. D., & Mandelbaum, A. (2009). Workload forecasting for a call center: Methodology and a case study. *Annals of Applied Statistics*, *3*, 1403–1447.

Antipov, A., & Meade, N. (2002). Forecasting call frequency at a financial services call centre. *Journal of the Operational Research Society*, *53*, 953–960.

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, *8*(1), 69–80.

Bianchi, L., Jarrett, J., & Hanumara, R. C. (1998). Improving forecasting for telemarketing centers by ARIMA modeling with intervention. *International Journal of Forecasting*, *14*, 497–504.

Broock, W. A., Scheinkman, J. A., Dechert, W. D., & LeBaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, *15*(3), 197–235.

Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H. P., Zeltyn, S., & Zhao, L. (2005). Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American Statistical Association*, *100*, 36–50.

Chatfield, C. (1985). Comments on 'exponential smoothing: The state of the art'. *Journal of Forecasting*, *4*, 30.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–73.

Crone, S. F., & Kourentzes, N. (2010). Feature selection for time series prediction – A combined filter and wrapper approach for neural networks. *Neurocomputing*, *73*, 1923–1936.

Dalrymple, D. J. (1978). Using Box-Jenkins techniques in sales forecasting. *Journal of Business Research*, *6*(2), 133–145.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*, 1513–1527.

Fildes, R., & Kumar, V. (2002). Telecommunications demand forecasting - A review. *International Journal of Forecasting*, *18*, 489–522.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3–23.

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, *5*, 79–141.

Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*, 1–28.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., & Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, *191*(1), 207–222.

Hamid, S. A., & Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 index futures prices. *Journal of Business Research*, *57*(10), 1116–1125.

Hyndman, R. (2010). Forecasting functions for time series. *Forecast. R package version 2.09 ed*.

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.

Ibrahim, R., & L'ecuyer, P. (2013). Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models. *Manufacturing & Service Operations Management*, *15*, 72–85.

Kourentzes, N., Barrow, D. K., & Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, *41*(9), 4235–4244.

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*. John Wiley & Sons.

Mandelbaum, A. (2006). *Call centers: Research bibliography with abstracts [online]*. Technion, Israel Institute of Technology (Available: http://iew3.technion.ac.il/serveng/References/ccbib.pdf [Accessed: 01–01-2014]).

Mandelbaum, A., Sakov, A., & Zeltyn, S. (2000). *Empirical analysis of a call center*. Israel Institute of Technology: Technion.

Millan-Ruiz, D., Pacheco, J., Hidalgo, J. I., & Velez, J. L. (2010). Forecasting in a multi-skill call centre. *Artificial intelligence and soft computing, Pt ii, 6114* (pp. 582–589).

Pacheco, J., Millan-Ruiz, D., & Velez, J. L. (2009). Neural networks for forecasting in a multi-skill call centre. *Engineering Applications of Neural Networks, Proceedings*, *43*, 291–300.

Petropoulos, F., & Kourentzes, N. (2014). Improving forecasting via multiple temporal aggregation. *Foresight: The International Journal of Applied Forecasting*, *2014*(34), 12–17.

Shen, H. P. (2009). On modeling and forecasting time series of smooth curves. *Technometrics*, *51*, 227–238.

Shen, H. P., & Huang, J. Z. (2005). Analysis of call centre arrival data using singular value decomposition. *Applied Stochastic Models in Business and Industry*, *21*, 251–263.

Shen, H. P., & Huang, J. Z. (2008a). Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management. *Annals of Applied Statistics*, *2*, 601–623.

Shen, H. P., & Huang, J. Z. (2008b). Interday forecasting and intraday updating of call center arrivals. *M&Som-Manufacturing & Service Operations Management*, *10*, 391–410.

Tandberg, D., Easom, L. J., & Qualls, C. (1995). Time-series forecasts of poison center call volume. *Journal of Toxicology, Clinical Toxicology*, *33*, 11–18.

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

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, *54*, 799–805.

Taylor, J. W. (2008a). A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Science*, *54*, 253–265.

Taylor, J. W. (2008b). Exponentially weighted information criteria for selecting among forecasting models. *International Journal of Forecasting*, *24*, 513–524.

Taylor, J. W. (2010). Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles. *International Journal of Forecasting*, *26*, 627–646.

Taylor, J. W. (2012). Density forecasting of intraday call center arrivals using models based on exponential smoothing. *Management Science*, *58*, 534–549.

Taylor, J. W., & Snyder, R. D. (2012). Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing. *Omega-International Journal of Management Science*, *40*, 748–757.

Temraz, H. K., Salama, M. M. A., & Chikhani, A. Y. (1997). Review of electric load forecasting methods. *1997 Canadian conference on electrical and computer engineering* (pp. 289–292).

Theodosiou, M. (2011). Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting*, *27*(4), 1178–1195.

Timmermann, A. (2006). Forecast combinations. In G. Elliot, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Amsterdam Elsevier).

Tych, W., Pedregal, D. J., Young, P. C., & Davies, J. (2002). An unobserved component model for multi-rate forecasting of telephone call demand: The design of a forecasting support system. *International Journal of Forecasting*, *18*, 673–695.

Weinberg, J., Brown, L. D., & Stroud, J. R. (2007). Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. *Journal of the American Statistical Association*, *102*, 1185–1198.

West, M. (1997). Time series decomposition. *Biometrika*, *84*(2), 489–494.

Willis, H. L., & Northcotegreen, J. E. D. (1983). Spatial electric-load forecasting - A tutorial review. *Proceedings of the IEEE*, *71*, 232–253.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.

Zhang, G. P., & Berardi, V. L. (2001). Time series forecasting with neural network ensembles: An application for exchange rate prediction. *Journal of the Operational Research Society*, *52*, 652–664.

Zhang, G. Q., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*, 35–62.