CALL CENTER PERFORMANCE WITH DIRECT RESPONSE ADVERTISING

Meltem Kiygi-Calli  
Marcel Weverbergh  
Philip Hans Franses

EI2017-04

Abstract

This study investigates the manpower planning and the performance of a national call center dealing with car repairs and on the road interventions. We model the impact of advertising on the capacity required. The starting point is a forecasting model for the incoming calls, where we take into account the impact of radio and TV commercials. An autoregressive-distributed lag model is used, which accounts for time-varying autoregressive effects. With our estimation results, we construct a forecasting tool based on a weekly media plan, and compare its forecasts with forecasts under baseline conditions without advertising. Next, the forecasts are fed into the capacity planning simulation module. We simulate the process directly at the level of seconds. This simulation mimics the service level requirements and queue behavior (waiting times, abandoned calls and idle time). The simulations show that the call center is operating at a high level of efficiency and performance. At the same time, we show that advertising may lead to a temporary overload of the system, and this increases the amount of abandoned calls, which is suboptimal.

Keywords: Call Center, Service Operations, Capacity Management, Discrete Event Simulation

---

1Faculty of Economics and Administrative Sciences, Okan University, Tuzla Campus, 34959, Istanbul, Turkey Email address: meltem.kiygicalli@okan.edu.tr; Tel: +90 216 677 16 30; Fax: +90 216 677 16 67.

2Faculty of Applied Economics, Marketing Department, University of Antwerp, Prinsstraat 13, B-2000 Antwerp, Belgium. Email address: marcel.weverbergh@ua.ac.be; Tel: +32 3 265 44 88; Fax: +32 3 265 47 99.

3 Econometric Institute, Erasmus University Rotterdam, POBox 1738, NL-3000 DR Rotterdam, The Nederlands. Email address: franses@ese.eur.nl; Tel: +31 10 408 12 73; Fax: +31 10 408 91 62.

The authors acknowledge the financial support of the Flanders Research Foundation (FWO) (Project number: G.0361.07).
1. Introduction

The effectiveness of call center operations is a crucial link in the marketing activity for companies organizing their business around a centralized, telephone-based collection of orders and ensuing planning of order executions. For example, management of the company we study in this paper indicated to us their difficulty of providing the desired service level in the call center at peak call moments. This is because these moments are short in time and staffing of the call center is not always sufficiently flexible to accommodate this.

In this study, we analyze the effectiveness of the operations of a specific call center, and we indicate routes for improvement. Our approach can be viewed as prototypical for companies in similar industries. The purpose of our analysis is to investigate to what extent the results from the advertising/calls relationship discussed in Kiygi-Calli, Weverbergh and Franses (2012) can be used in the capacity or manpower planning of the call center. The key research question is to what extent the call center capacity can cope with the increase in calls caused by advertising. We thus investigate whether advertising affects the efficiency of the call center.

The operation of a call center relates to queuing theory. For simpler, stable queuing systems, analytical relationships can be used. In our case, the average number of servers or agents varies from 1 to 48 or more over the course of a day, and the number of incoming calls is characterized by a complex dynamic process. Therefore, we resort to simulation of the system rather than to analytical methods. Two measures of the call center activity are of interest for our company. The first one deals with so-called ‘relevant calls’. These are calls pertaining to the business as such, see Kiygi-Calli, Weverbergh and Franses (2012) for more details. The second measure starts from all incoming calls, whether related to information requests or orders or not. The second measure is more relevant for the operation of the call center, while the first one is closer to the actual conversion of calls to sales. We use a discrete event simulation that models the process directly, and which allows for an integration of the forecasting system and the call center simulation.

The paper is organized as follows. Section 1 describes the call center operation and the objectives the management wants to achieve in terms of effectiveness. In Section 2, we briefly discuss the literature related to our topic. Section 3 and 4 explain the data and the modeling of the incoming
calls, followed by results of the model in Section 5. Section 6 contains the simulation model, and
Section 7 discusses the simulation results and in Section 8 we discuss the conclusions.

2. Background

Call centers where customers’ questions, needs and requests are handled are important parts of
many large organizations. In these centers the traditional telephone service is enhanced by
additional customer contact channels, for instance by Interactive Voice Response (IVR) (Koole
and Mandelbaum, 2002). Most companies communicate with their customers via internally-
managed or outsourced call centers. There are studies on call center capacity management, which
focus on performance as measured by waiting times (see for example Aksin, Armony and
Mehrotra, 2007).

The random variation in incoming calls of a call center is very important when assessing call center
performance (Betts, Meadows and Walley, 2000). According to the call center managers’ needs,
the call center data might be stored in real time or aggregated over time to some extent with the
purpose of measuring the performance of the call center. In the evaluation of the call center
performance, typical service level metrics are the waiting times for callers, the number of
abandoned calls due to excessive waiting times and the amount of time, call center operators are
idle (there is excess capacity). Discrete event simulation systems are driven by events in a time
listing of events and the byproducts from that chronology (Allen, 2001). These simulation systems
are discussed in depth by Cassandas and Lafortune (2008) and Wainer (2009), among others.

The main instrument for managing the call center is the number of agents and their working
schedules in relation to the volume of calls. A balance has to be found between excess capacity
(idle time of operators) and experience in dealing with calls. Conceptually, call centers are single
queue first-in-first-out, multi-server queuing systems, which are one of the standard topics in the
operations research literature (Hillier and Lieberman, 2005).

Kolesar and Green (1998) apply a queuing theory based approach in the context of call center
capacity management. They conclude that, in order to achieve a high service level, more staff is
required, which is costly for the companies. The disadvantages of queuing models may be that
they tend to be time-consuming, complex, simplifying reality and unreliable (Buist, Chan and

Cezik and L’Ecuyer (2008) optimize the staffing and the scheduling problems and use an iterative cutting plane method which relies on the service level function which is concave in the number of servers for minimizing staffing costs in a call center. Atlason, Epelman, and Henderson (2004) also use the iterative cutting plane method in their study. In Avramidis et al. (2009) the objective of the study is also to minimize the total cost of the agents.

Modeling of a call center needs to be performed based on a detailed analysis of the operational data. The peaks in incoming calls need to be understood and the factors which make the difference in the number of incoming calls should be taken into account for different call centers (Soyer and Tarimcilar, 2007).

Aksin, Armony and Mehrotra (2007) conduct a survey study on call center operations management and review recent call center research. They conclude that forecasting models play an important role in operations and are a critical input for resource acquisition. Therefore, companies should take into account their marketing activities in order to forecast incoming calls to the call center. Furthermore, companies need to develop a balance between the operations and marketing activities (Kotler, 1991). The marketing department, setting the marketing policies, plans advertising schedules of the company, and the operations should be arranged in such a way that they do not result in conflicts due to inconsistent objectives (Eliashberg and Steinberg, 1993).

In this study, we investigate the impact of the calls volume and the operational efficiency, where the calls volume is affected by direct response commercials used by the advertising company. In the first part of the study, we investigate the impact of direct response commercials on incoming calls at the national call center as moderated by different media (radio and TV channels) and time of broadcasting. We use an autoregressive-distributed lag model, with total incoming calls aggregated at 15-minute intervals as the dependent variable. Using the estimation results of the model, we construct a simulation tool based on a weekly media plan. It computes the baseline forecasts (without advertising) and the forecasts resulting from a media plan relying on average
Gross Rating Points (GRPs) by channel and time slot (quarter of an hour). GRPs is a measure of audience size and it is calculated as “reach*average frequency” (Govoni, 2004). In the second part of the study, we link the forecasts to the capacity planning. This requires a ‘disaggregation’ from the forecasting time intervals (quarter of an hour) to the interval with seconds. This gives the arrival rate for incoming calls at each moment of the planning week.

3. Data

We analyze the data related to a national call center, which collects all requests from consumers. The company broadcasts direct-response radio and TV commercials on national channels. In the first part of the study, we measure the impact of direct-response commercials on the number of incoming calls. The call center operates on a 24/7 basis and the number of incoming calls is recorded in real time (seconds). Service centers are located in two different regions. These are Flanders (the Dutch-speaking part of Belgium) and Wallonia (the French-speaking part of Belgium). Although the media plans and communication channels are different in each region, a single telephone number is used and agents, who all are bilingual, handle incoming calls irrespective of the language choice of the customers. Calls arrive at the central office through two channels. There are incoming 0800 calls, and these are calls directly to the call center number advertised through commercials or on the website of the company. There are also calls redirected from regional service centers. The data cover the period from June 28, 2008 to December 16, 2009. The data are provided to us for 15-minute intervals, resulting in 51552 observations. Incoming calls are registered as Dutch or French calls, according to the language choice of the customers. In addition to these, there are also undefined calls, that is, calls for which no language registration is available. Approximately 28% of the total calls are undefined. We treat them as a third region.

Figure 1 shows the generation of incoming calls by region. In total, there are 3358 radio commercials recorded, spread over 14 radio stations broadcasted between 6 AM to 8 PM and the number of TV commercials is 3100 broadcasted on 9 TV stations between 11 AM and midnight. The flow chart of the call center process is sketched in Figure 2. In the data, service times are given and incoming calls are defined as handled by agent or abandoned by the caller.

--Insert Figure 1 about here--

--Insert Figure 2 about here--
The maximum seat capacity of the call center is 61 and the total number of potentially available agents is 67. During the night hours, only one agent handles the incoming calls. During the peak hours, the number of agents might reach 48 in total. In the interactive voice response (IVR) system customers are asked to choose a language, Dutch or French. After the language choice, customers are asked to choose the type of service. Type of the service can be either repair or replacement. The customers calling for replacement might ask detailed questions and need more information. Because of that, the call durations of this type of calls is longer than the normal duration for repair calls. In order to answer the replacement calls, agents need to receive special training. There is also an outsourced call center operating between 8 AM and 8 PM on weekdays. Calls related to repairs are redirected to this external call center automatically during the mentioned hours. Customers whose waiting times are longer than 1 minute are also redirected to the external call center. The target service level of the call center is handling 80% of the calls with a 12-second waiting time limit.

4. Calls Forecasting Model

The call center operates according to a single queue/multi-server principle, and the total number of the incoming calls will be used as a dependent variable in our model. For each of the three regions (Dutch speaking, French speaking and undefined) we estimate a forecasting function based on an Autoregressive Distributed Lag (ADL) model. These models shall be used for forecasting the number of calls over a planning horizon of a week. Kiygi-Calli, Weverbergh and Franses (2012) use a Linear Mixed Model (LMM) for similar kinds of data, but here we do not adopt the LMM model for the following reasons:

a) The observation period is shorter, providing less information about individual time slots.
b) We use fifteen-minute intervals rather than hourly data, as the purpose in the current paper is on detailed forecasting of the call center dynamics, rather than on assessing the impact of advertising.
c) The increased frequency leads to less advertising spots per slot (and more slots without advertising), limiting the potential of LMM models.
d) While the LMM model is a powerful vehicle to test for heterogeneity of advertising effects over channels, time slots and spot characteristics, it is not necessarily superior in terms of forecasting (Kiygi-Calli, Weverbergh and Franses, 2017).

e) We take into account the essential conclusions obtained from the hourly LMM models. These are

a. Advertising effects are not very heterogeneous. For forecasting, the overall GRP levels for radio and TV are adequate as predictors.

b. The autoregressive structure shows a goniometric pattern over the day.

The result of these five considerations leads to a model formulation, which is close to the LMM model, but imposes more homogeneity on the advertising effects. The most important difference is that we do not include any random effects., making the model more easy to handle.

We determine the best lag structures for the AR terms and GRP effects by means of Almon lags based on the significance and the pattern of the estimates for increasing lags. In a first step, the AR process is modeled as an Almon lag of order (10,10), equivalent to 10 lags without restriction on the lag coefficients. The values of the AR coefficients obtained in this way are presented in Figure 3. It is clearly observed that the values of AR coefficients are decreasing with the order of lags. Because the pattern changes at lag 7 and the lags with higher degrees are insignificant, the order of the AR term is decided to be 7 resulting in AR effects for lags up to 7 quarters of hours. From a visual inspection (see Figure 3), it can also be concluded that the AR term follows a quadratic reaction pattern. We thus conclude that Almon lags of order 7 and degree 2 provide a good representation of the structure of the AR term for the model. In addition to the intraday AR effects, it is found that the daily and weekly AR terms are also significant.

--Insert Figure 3 about here--

Figure 4 presents the pattern of radio GRP coefficients of order 6 and degree 6. The radio GRP effects are significant up to 4 quarters of hours and up to second degree. From these radio GRPs, we define the DL pattern by means of Almon lags of order 4 and degree 2 (4,2).

--Insert Figure 4 about here--
The radio and TV GRP effects display a homogeneous pattern in Kiygi-Calli, Weverbergh and Franses (2012), and therefore, we do not model these effects as heterogeneous.

The final model specification is given by

\[
Y_{R,t} = \mu + (\lambda_1 + \theta_1 \sin \left( \frac{2\pi q d}{96} \right) + \theta_2 \cos \left( \frac{2\pi q d}{96} \right))Y_{R,t-1} + (\lambda_2 + \theta_3 \sin \left( \frac{2\pi q d}{96} \right) + \theta_4 \cos \left( \frac{2\pi q d}{96} \right))Y_{R,t-2} \\
+ (\lambda_3 + \theta_5 \sin \left( \frac{2\pi q d}{96} \right) + \theta_6 \cos \left( \frac{2\pi q d}{96} \right))Y_{R,t-3} + \lambda_4 Y_{R,t-4} + \lambda_5 Y_{R,t-5} \\
+ \lambda_6 Y_{R,t-6} + \lambda_7 Y_{R,t-7} + \lambda_8 Y_{R,t-96} + \lambda_9 Y_{R,t-672} \\
+ \delta^R \sin \left( \frac{2\pi t}{672 \times 52} \right) + \delta^S \cos \left( \frac{2\pi t}{672 \times 52} \right) + \beta_1 T + \beta_2 B_t + \beta_3 B_{t-96} + \beta_4 B_{t+96} \\
+ \phi^0 R_{R,t} + \phi^1 R_{R,t-1} + \phi^2 R_{R,t-2} + \phi^3 R_{R,t-3} + \phi^4 R_{R,t-4} + \phi^5 R_{R,t-5} \\
+ \phi^6 R_{R,t-6} + \phi^7 R_{R,t-7} + \sum_{hd=1}^{23} \phi^h_{R,t} D_{t,hd} RD_{d-1,R,t} \\
+ \gamma^0 TV_{R,t} + \gamma^1 TV_{R,t-1} + \sum_{hd=1}^{23} \gamma^h_{R,t} D_{t,hd} TVD_{d-1,R,t} \\
+ \sum_{q=1}^{95} \phi^q D_{i,q} + \sum_{d=1}^{6} \phi^d D_{i,d} + \epsilon_t
\]

(1)

Where

\(Y_{R,t} = \log(Calls_{R,t} + 1)\)

\(R_{R,t} = \log(RadioGRP_{R,t} + 1)\)
\[ TV_{R,t} = \log(TVGRP_{R,t} + 1) \]

\[ \sin\left(\frac{2\pi}{672 \times 52}\right), \cos\left(\frac{2\pi}{672 \times 52}\right) \] are the harmonic or goniometric regressors, capturing the intra-year seasonality in the data,

\( qd=1,\ldots,96 \) denotes the \( qd \)’th quarter hour of the day. \( \sin\left(\frac{2\pi qd}{96}\right), \cos\left(\frac{2\pi qd}{96}\right) \) are the harmonic or goniometric regressors, capturing the intra-day seasonality in the data,

\( Tr_t \) is a trend defined as \( \frac{t}{672 \times 52} \).

\( B_t \) is a dummy variable for bank holidays,

Advertising terms: \( RD_{d-1,R,t} \) and \( TVD_{d-1,R,t} \) denotes the log of the total amount of Radio GRPs and TV GRPs during the previous day and \( D_{t,hd} \) is a dummy variable for hour \( hd \) of a day while \( D_{t,q} \) is a dummy variable for quarter \( q \) in the week and \( D_{t,d} \) is a dummy variable for day \( d \) in the week.

In the model, calls and GRPs are in logarithms after augmentation with 1. Kiygi-Calli, Weverbergh and Franses (2017) show that for low levels of aggregation a logarithmic transformation is preferable. The first three order autoregressive terms exhibit goniometric behavior, which is also reported in Kiygi-Calli, Weverbergh and Franses (2012).

5. Results of the Forecasting Model

Table 1 shows the parameter estimates of model (1). The regions are Belgium-North (Region 1), Belgium-South (Region 2) and Undefined Calls (Region 3). Estimation results of the three regions clearly present similarities in terms of the significant lag orders. According to the estimation results, we find a highly significant goniometric wave for the autoregressive (AR) effects. The first three lags of the AR terms show a pronounced and highly significant goniometric pattern. They are specified as a wave with a periodicity of one day. For the first region, Figure 5 presents the waves of first, second and third order goniometric AR terms by quarter of an hour for one week. Besides these AR terms, the total AR term which is the sum of the goniometric, daily and weekly
AR terms is also shown in Figure 5. The total autoregressive term has a peak of 1 at 11 AM and a low of 0.5 at 11 PM for each day for region 1.

The one-day-lagged (AR(96)) terms have an effect of 0.025, 0.020, and 0.017 for the three regions, respectively. Furthermore, the one-week-lagged (AR(672)) terms are highly significant with estimates of 0.101, 0.086, and 0.087 for the three regions.

Highly significant yearly seasonality is also found for three regions. Figure 1 suggests that there is a goniometric yearly cycle in the three regions. In the first region (Dutch speaking) we found a negative trend with a value of 0.009. During bank holidays, the number of incoming calls decreases by approximately 10-12% for each region.

For the three regions, we found highly significant and positive advertising effects during the current time slot for radio and TV. The results of lagged radio effects are given in Table 1. Radio GRPs effects are significant for lags up to 7 quarters of an hour. For the first region, one-quarter of an hour lagged of TV GRPS is also significant with a value of 0.041. The effects of previous day radio GRPs and TV GRPs are also significant.

The coefficient of determination ($R^2$) for the forecasting models are 0.88 for Region 1, 0.86 for Region 2 and 0.82 for Region 3. Higher orders of lags usually increase the $R^2$. However, this is not observed in the results of our final model showing that the $R^2$ values are quite stable.

The in-sample Root Mean Square Errors (RMSE) for the logarithms of calls are 0.398, 0.447 and 0.423 for three regions. The white noise tests (Barlett’s Kolmogorov-Smirnov Statistic) are 0.014, 0.015, and 0.013 with p-values 0.0002 or smaller than 0.0001. A small amount of autocorrelation apparently exists in the residuals of the model. The maximum value of autocorrelation in the residuals is 0.015. Autocorrelations of this magnitude however do not affect the results in a substantial way.

Figure 6 shows the forecasts of the model for benchmark call levels for an inactive week (without any advertising broadcasted) and the week-ahead forecasts for the week with a media plan which
is called active week. In the data, an active week is typically preceded by a week without ads. The difference between the benchmark call level and the week-ahead forecast gives the impact of the commercials. For generating week-ahead forecasts, we use a real media plan, which is planned for a week by the company. The model can be used to investigate the impact on the next week, however our focus is to investigate the impact during the active week.

6. Simulation

In the second part of this study, we evaluate the call center performance of the company when direct-response commercials are broadcasted. The call centers are dynamic systems that can only be modeled by a simulation. In the simulation, we link the week-ahead forecast of the forecasting model, which is discussed above, with the call center capacity planning system. Therefore, the total of the three regions’ simulated forecast calls is one of the input of the simulation system. We evaluate the results of the simulation by means of manpower planning policy of the company during the inactive and the active weeks. An active week means that the company broadcasts direct-response commercials, which are spread throughout the week at irregularly spaced intervals. In contrast, an inactive week means that there is no advertising broadcasted during the entire week.

We calculate the residual correlation by means of Pearson Correlation Coefficient for the 3 regions. The statistics show that the residuals of the estimates are significantly correlated (p < .0001). Table 2 gives the covariance matrix of the three forecast models’ residuals. As the residuals are correlated, the sum of the residuals’ variances is calculated as

\[ V = u' E(\hat{e}_R' e_R) u \]

where

\( \hat{e}_R \) is the matrix of residuals by region, \( u \) is the sum vector and equal to \( \{1,1,1\} \), \( V \) is the covariance matrix of the residuals for 3 regions. We find the total variance as 0.65 and thus take the forecasting error of the arrivals as a normal distribution with a mean of zero and a variance of 0.65. In order to calculate the total forecast, we draw a random disturbance from the normal distribution (0, 0.65).
and add the back-transformed random disturbance to the simulated forecast calls. Subsequently, we use the calculated total forecast as an input of the call center simulation.

---Insert Table 2 about here---

For the inactive week, we forecast the benchmark call level for a week without advertising broadcasted. For the active week, we forecast the incoming calls according to a typical media plan. Figure 7 represents the average number of agents for non-holidays and the number of agents in active advertising weeks. From the figure, we can conclude that for the active weeks the company changes capacity planning policy, that is, the number of agents is increased in active weeks.

---Insert Figure 7 about here---

In order to simulate the manpower planning system in seconds, we generate a discrete event simulation system. The simulation system is explained in the “Discrete Event Simulation” section in detail.

Our econometric model leads to high demands in terms of horizon (a week), time unit, and complexity of the call process (lags up to seven quarters of an hour, completed by day and week lags). Also, the call center complexity is quite high (from 1 to 48 agents during a day cycle). In the forecasting model discussed above, total calls are recorded at 15-minute intervals, rather than for hours as used in Kiygi-Calli, Weyerbergh and Franses (2012). We take into account the media plans and reach (GRP’s) of the advertisements. The inputs into this forecasting system are:

a. Last week calls. It is assumed that last week calls are available before the planning week. If not, the forecasting process can be based on expected call levels for last week, derived from the econometric models for North, South and undetermined calls.

b. the media plan for the planning week

The results of this forecasting process are imported into the simulation, together with the manpower schedule for the planning week. In the simulation process, there are two inputs and these are arrivals and manpower. The calculation of the active week arrivals is explained above. In the active week, the available manpower is the average number of agents observed during each 15 minutes interval of the week analyzed. In the inactive week, manpower is modeled as the
average manpower over all weeks without advertising by 15 minutes intervals. These averages are computed excluding holidays.

During inactive weeks the average number of agents in the call center changes from 1 during the night to 35 at peak hours. For active weeks, manpower at peak times increases to 48 agents.

In the simulation, incoming calls go into a queue. Customers may terminate the call before an agent responds to the call. These are called as “Abandoned Calls”. Figure 8 shows the relationship between the number of abandoned calls and the average of incoming calls per agent by quarter of a week. From this figure, we can conclude that there is a threshold at a ratio of 1.5 incoming calls per agent. When the ratio of incoming calls per agent is less than 1.5, 12% of the incoming calls is abandoned. When the ratio is greater than 1.5, the abandoned calls function changes.

As discussed in the data section, the external call center operates between 8 AM and 8 PM. Between these hours the incoming calls, which are related to repair services, are directed to the external call center. The repair calls to the external call center are 7% of the incoming calls during daytime operations.

During daytime operations, customers who are waiting more than 90 seconds are redirected to the external call center. From the data, 5.5% of the total calls are overflows with an associated average waiting time of 196 seconds. Figure 9 shows the relationship between the incoming calls per agent and redirected overflows between 8 AM and 8 PM. From the figure, we can conclude that there is a threshold at the incoming calls per manpower ratio of 1.5. If during the daytime the ratio of incoming calls per agent is less than 1.5, approximately 1% of the incoming calls are redirected to the external call center. If the ratio is greater than 1.5, then the function changes.

Service time is the duration of the conversation between the customer and the agent. Subtracting abandoned, overflow and redirected calls from the total incoming calls gives the net incoming calls, which are effectively processed by the internal call center. The number of calls that agents can handle is linked to the available capacity. The waiting time is the amount of time spent in the queue.
Discrete Event Simulation

The managerial objectives are stated as: 80 percent of customers should have a waiting time smaller than 12 seconds. Standard mathematical software, such as Mathematica, Matlab, and Maple, allow a discrete event simulation in a straightforward way. We can make use of the sparse nature of the process by only tracking the calls, at the relevant interarrival times, and simulating the system is more efficient by an order of magnitude and more flexible for dealing with a varying number of agents.

The simulation algorithm goes as follows:

- Compute the week-ahead calls process as described above.
- Simulate the process at a per second level for a week.
- For each second the number of calls is drawn from a Poisson distribution, with a mean arrival rate per second $\lambda_s$, obtained by means of a linear interpolation between $\lambda_{qd}/(60*15)$ and $\lambda_{qd+1}/(60*15)$ where $\lambda_{qd}$ is the number of calls obtained from the forecasts during the relevant 15-minute interval. This results in a simulation of received calls, which yields about 200 calls more than the average total for a week in the data set. This is in line with expectations, because the observed average decreases somewhat during holidays.

---Insert Figure 10 about here---

---Insert Figure 11 about here---

- For all arrivals generated at a particular second, we do the following

  i. Each arrival is equipped with a random service time, drawn from the service time distribution (see Figure 10), and a random ‘patience’, which tells the system after how much time a call will be abandoned without service, drawn from the ‘abandoned calls’ distribution (see Figure 11). Figure 10 and Figure 11 show the distributions of the service times and the duration of the abandoned calls in the raw data. Service duration is defined as a log normal distribution with a mean of 221 (exponential of 5.4) seconds.
and a standard deviation of 2.12 (exponential of 0.75) seconds when there is only one server in the system. We define the duration of an abandoned call as the patience of the customer. Figure 11 shows that there is a threshold at 5.47 (exponential of 1.7) seconds for patience. Between 5.47 (exponential of 1.7) and 148 (exponential of 5) seconds the percentage of the patience is approximately uniform. The mean of patience is 31.2 (exponential of 3.44) seconds with the standard deviation of 2.56 (exponential of 0.94) seconds. The average service time observed is 3.68 minutes, which is used in the discrete event simulation. Agents perform some administrative task between calls, but there are no hard data on this. The discrete event simulation does not take into account the time spent between calls.

ii. Determine the available number of agents at that point in time, and read the earliest time an agent will be available.

iii. Calculate the necessary waiting time, compare to patience. If patience is longer than waiting time, update the time the earliest available agent will be free again.

iv. update metrics: total arrivals, abandoned, waiting times, processing time, redirected calls to the external call center, overflows and idle capacity in the system

• compute overall results: number of calls processed, and waiting times distribution

7. Results of the Simulation

The week-ahead forecasts are computed for inactive and active weeks and the waiting times of the customers are evaluated for both weeks. Table 3 gives the percentage of the waiting times for the active and inactive weeks. For the active week the percentage of waiting times less than 5 seconds is 74.87%, the percentage of waiting times less than 12 seconds is 81.45% and the percentage of waiting times less than 20 sec is 100%.

--Insert Table 3 about here--

In contrast, for the inactive week the percentage of waiting times less than 5 seconds is 82.24%, the percentage of waiting times less than 12 seconds is 87.38% and the percentage of waiting times less than 20 seconds is 100%.
Figure 12 gives the histogram of the waiting times. It shows that the waiting times in the active weeks are slightly longer than the waiting times in the inactive weeks. In the active weeks, we find that 4.12% of the callers leave the system within 10 seconds. On the other hand, in the inactive weeks, the percentage of leaving the system within 10 seconds decreases to 3.06.

Because the waiting times are less than 60 seconds, there are no overflows to the external call center. Table 4 represents the percentage of the abandoned calls. For the active week, we find the abandoned calls as 16% of the total calls and 54.5% of the abandoned calls occur between 6 AM and 12 AM. On the other hand for the inactive week the percentage of abandoned calls is 11% of the total calls and 49% of the abandoned calls occur between 6 AM to 12 AM.

We also calculate the Idle Capacity of the system.

On average, the idle capacity is 44% and 57% of the total capacity in inactive and active week, respectively. Figure 13 shows the idle capacity in minutes. The idle capacity of an active week is higher than the idle capacity of an inactive week.

Figure 14 (a) and (b) show the number of people waiting in the queue for inactive and active weeks, which are obtained from discrete event simulation. The queue increases in active weeks which results in more abandoned calls. Since advertising effectiveness is measured by means of (relevant) incoming calls answered, this increase in abandoned calls is a potential threat to the advertising effectiveness.

8. Conclusions

In the first part of this study, we constructed a forecasting model for incoming calls taking into account the impact of direct response commercials. We used an ADL model and found highly significant intra-day pattern for the first three lags of the autoregressive terms. The forecasts for this model were used as inputs for an evaluation of the manpower planning of the call center. The
stated objectives for the call center are quite ambitious: for 80 percent of incoming calls, waiting times should be less than 12 seconds. We found that this objective is not fully realized.

Evaluation is done by the discrete event simulation. The ad-hoc discrete event simulation deals with the issues in an efficient way, and allows to assess the performance of the system in great detail. The results of the system dynamics simulation show that the idle capacity in the active week is higher than the one in the inactive week. In addition to this, we find that the waiting times are also higher in the active week. This tradeoff shows the difficulty for balancing between excess capacity and waiting times.

The inputs of the simulation are arrivals and manpower. The discrete event simulation uses random service times and patience based on the available data. The waiting times applied in the discrete simulation are based on recorded duration.

Apparently, the firm anticipates well to the increase in calls during active weeks, as witnessed by the ratio incoming calls per agent. The simulation shows increases in queue size, waiting times and abandoned calls in active weeks. On the positive side, the indications from our simulation are that the adjustments in manpower between active weeks and inactive weeks have resulted in an improved (on average 6 calls/agent in an hour) ratio of calls to agents. On the down side, for active weeks, the system is inherently less manageable, as indicated by a simultaneous increase in idle capacity and waiting times.

The system operates close to the performance levels stated in the managerial objectives. This notwithstanding that the efficiency of the system is lower during active weeks, which shows that advertising has a detrimental effect on peak calls, leading to higher waiting times and abandoned calls. According to the managerial feedback, manpower planning is less flexible during the weekend. The simulation also shows that much of the increased manpower capacity results in an increase in idle time.

**Managerial implications**

In order to answer the question to what extent the evaluation developed here can be implemented as a capacity planning tool, some further issues should be resolved:
a) Manpower planning

Is the manpower planning closed loop (this means can it be adjusted on the spot on the basis of the number of calls arriving or waiting time) or is it open loop as assumed in our simulation (with predefined capacities for each time slot for a week)?

b) Data

Ideally, some additional data should be estimated or measured. One such item would be a distinction between waiting time and service time for the calls. In this study, we have the total time that a customer spends in the system.

c) Validity checks

A major point is to get a grip on realistic service times: the micro-simulation is too optimistic in terms of waiting times, and is too positive with respect to the performance of the system, mainly because better understanding of administrative time required between handling calls is required.

d) Service strategy

From a managerial point of view, the task is to reduce the demands on the call center to the extent possible. This is attempted by presenting the internet platform and the call center side by side in the advertisements. However, the call center remains crucial because the advertiser relies on impulse calling behavior from the target customers.

The main contributions of the study are as follows: This study demonstrates the importance of the link between marketing activities and operational management. The forecasting model developed in this study shows that the autoregressive (AR) and distributed lag (DL) structures are in line with the AR and DL results obtained in the hourly models discussed in Kiygi-Calli, Weverbergh and Franses (2012). The seasonal heterogeneous structure becomes more complex for data at 15-minute intervals. However, the model fit remains roughly 90%, close to what is obtained at hourly data intervals (Kiygi-Calli, Weverbergh and Franses, 2012). In this study, the number of incoming calls is used as a dependent variable. In Kiygi-Calli, Weverbergh and Franses (2012), the number of relevant 0800 calls is used as an independent variable of the advertising response model. As the
dependent variables of the models are different in both studies the advertising impact results obtained in both are not directly comparable.

For further research, alternative aggregation levels of the data can be analyzed in order to have the most effective forecasting system. This study can be conducted with a different periodicity of the estimation, which is shorter than 15 minutes intervals.

References


Figure 1 Calls generation
Figure 2 Overview of the call center process (Service 1: Replacement Calls, Service 2: Repair)
Figure 3 Almon lag estimates of the autoregressive term
Figure 4 Almon lag estimates of the radio GRPs
Figure 5 Goniometric AR terms for a week (Region 1)
Figure 6 Comparison of forecasts for active and inactive weeks
Figure 7 Comparison number of agents for active and inactive weeks
Figure 8 Abandoned calls versus incoming calls per agent
Figure 9 Overflow calls versus incoming calls per agent
Figure 10 Service duration in logarithm of seconds
Figure 11 Duration of abandoned calls in logarithm of seconds
Figure 12 Waiting times in seconds (Discrete Event Simulation) for inactive (a) and active (b) weeks
Figure 13 Idle capacity in minutes for inactive and active weeks
Figure 14 Queue size for inactive (a) and active (b) weeks
### Estimation results of equation (1)

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Parameter</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
<th>Parameter Name</th>
<th>Parameter</th>
<th>Region 1</th>
<th>Region 2</th>
<th>Region 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>µ</td>
<td>-0.011</td>
<td>0.078</td>
<td>0.034</td>
<td>Current Radio GRP</td>
<td>φ₀</td>
<td>0.049</td>
<td>0.044</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.019)</td>
<td>(0.021)</td>
<td>(0.020)</td>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>First-order lag</td>
<td>λ₁</td>
<td>0.217</td>
<td>0.230</td>
<td>0.171</td>
<td>1 qhr lag</td>
<td>φ₁</td>
<td>0.030</td>
<td>0.032</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>θ₁</td>
<td>0.034</td>
<td>0.038</td>
<td>0.048</td>
<td>2 qhr lag</td>
<td>φ₂</td>
<td>0.017</td>
<td>0.022</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>θ₂</td>
<td>-0.111</td>
<td>-0.091</td>
<td>-0.085</td>
<td>3 qhr lag</td>
<td>φ₃</td>
<td>0.010</td>
<td>0.015</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>Second-order lag</td>
<td>λ₂</td>
<td>0.152</td>
<td>0.156</td>
<td>0.129</td>
<td>4 qhr lag</td>
<td>φ₄</td>
<td>0.009</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>θ₃</td>
<td>0.012</td>
<td>0.017</td>
<td>0.019</td>
<td>5 qhr lag</td>
<td>φ₅</td>
<td>0</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>θ₄</td>
<td>-0.081</td>
<td>-0.074</td>
<td>-0.074</td>
<td>6 qhr lag</td>
<td>φ₆</td>
<td>0</td>
<td>0</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td></td>
<td>(0.004)</td>
<td></td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>Third-order lag</td>
<td>λ₃</td>
<td>0.100</td>
<td>0.099</td>
<td>0.095</td>
<td>7 qhr lag</td>
<td>φ₇</td>
<td>0</td>
<td>0</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td>(0.005)</td>
<td></td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>θ₅</td>
<td>-0.010</td>
<td>-0.004</td>
<td>-0.010</td>
<td>1 day lag</td>
<td>Hourly</td>
<td>0.084</td>
<td>0.060</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td>Effect</td>
<td>(0.010)</td>
<td>(0.008)</td>
<td>(0.011)</td>
</tr>
<tr>
<td></td>
<td>θ₆</td>
<td>-0.052</td>
<td>-0.057</td>
<td>-0.063</td>
<td>1 day lag</td>
<td>Hourly</td>
<td>0.041</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td></td>
<td>Effect</td>
<td>(0.010)</td>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>Fourth-order lag</td>
<td>λ₄</td>
<td>0.063</td>
<td>0.058</td>
<td>0.069</td>
<td>1 qhr lag</td>
<td>γ₀</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>λ₅</td>
<td>0.039</td>
<td>0.034</td>
<td>0.052</td>
<td>1 day lag</td>
<td>Hourly</td>
<td>0.041</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td>Effect</td>
<td>(0.010)</td>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>Sixth-order lag</td>
<td>λ₆</td>
<td>0.029</td>
<td>0.027</td>
<td>0.042</td>
<td>Seasonality</td>
<td>δₛ</td>
<td>-0.022</td>
<td>-0.030</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Seventh-order lag</td>
<td>λ₇</td>
<td>0.033</td>
<td>0.036</td>
<td>0.041</td>
<td>Time</td>
<td>δₖ</td>
<td>0.010</td>
<td>0.003</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>One-day lag</td>
<td>λ₉</td>
<td>0.025</td>
<td>0.020</td>
<td>0.017</td>
<td>Trend</td>
<td>β₁</td>
<td>-0.009</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
<td>(0.004)</td>
<td></td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>One-week lag</td>
<td>λ₆₇₂</td>
<td>0.101</td>
<td>0.086</td>
<td>0.087</td>
<td>Holiday</td>
<td>β₂</td>
<td>-0.118</td>
<td>-0.106</td>
<td>-0.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
</tbody>
</table>
Table 2 Covariance matrix of the three forecast models

<table>
<thead>
<tr>
<th>Covariance Matrix</th>
<th>Residuals (Region 1)</th>
<th>Residuals (Region 2)</th>
<th>Residuals (Region 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals (Region 1)</td>
<td>0.16</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Residuals (Region 2)</td>
<td>0.02</td>
<td>0.20</td>
<td>0.02</td>
</tr>
<tr>
<td>Residuals (Region 3)</td>
<td>0.02</td>
<td>0.02</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Table 3 Percentage of the waiting time comparison

<table>
<thead>
<tr>
<th>Waiting times</th>
<th>&lt; 5 seconds</th>
<th>&lt; 12 seconds</th>
<th>&lt; 20 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Week</td>
<td>74.87%</td>
<td>81.45%</td>
<td>100%</td>
</tr>
<tr>
<td>Inactive Week</td>
<td>82.24%</td>
<td>87.38%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 4 Percentage of the abandoned calls comparison

<table>
<thead>
<tr>
<th>Abandoned Calls</th>
<th>Active Week</th>
<th>Inactive Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of the abandoned calls in total</td>
<td>16%</td>
<td>11%</td>
</tr>
<tr>
<td>% of the abandoned calls in total (between 6 AM and 6 PM)</td>
<td>14.6%</td>
<td>9.7%</td>
</tr>
<tr>
<td>% of the abandoned calls between 6 AM and 12 AM in total abandoned calls</td>
<td>54.5%</td>
<td>49.2%</td>
</tr>
</tbody>
</table>