Modelling calls and effects

"All our agents are busy" is a phrase call centre customers dread. But basic statistical models can help customer service managers ensure staff supply meets caller demand. **Andrew Zelin** explains

Call centres represent a significant growth area of the world's economy and, for most of the largest organisations, are used as the mainstay for handling the majority of their customer calls. Call centres have been called the "factories of the 21st century" (bbc.in/1OwP4SC), and, like factories, owners of call centres need to be able to accurately and reliably forecast how much work they will have on any given day, so that the right number of staff can be at their desks at the right times to take calls.

Too few call operators will lead to poor customer experience, with increased waiting times, abandonment of calls and a backlog of frustrated customers. Too many staff members, by contrast, will lead to excessive operational costs and a negative impact on profitability. Either way, both customers and the organisation will suffer, and so will the staff, be it through stress-related burnout or boredom.

If one can predict call volumes sufficiently far into the future, then staff recruitment patterns can be accelerated or slowed down based on forecast need. How can one predict the magic figure of the number of calls that will come in on day X of week Y and month Z, between 10 a.m. and 11 a.m., so as to ensure call demand meets staffing supply? If you are a numerate planner with basic statistical knowledge, this article explains how you might generate forecasts with reasonable accuracy, using scientific, transparent, reproducible and replicable methods.

Note that the approach described is not restricted to call centres. You may be running services to process welfare benefit claims or dealing with telephone queries from residents about their housing. Indeed, it could cover any area of public or private sector service where demand needs to be predicted accurately: footfall in shops, crime/incident data, social care, refuse collection or even share prices and climate change. But for the purposes of this article, our example is a private medical insurance company where members call in requiring authorisation for medical treatment.

Thinking about forecasting

If one wants to accurately and reliably predict the number of calls that will need to be taken at a call centre at any given point in time, the best guide to the future is to look at what has happened in the past.

A good place to start is with the number of calls received on a typical day – the straight mean of all the days in your historical series. Then you can ask:

- Are there differences between what is typical for a Monday, a Tuesday, etc.?
- If you are open on public holidays, what is typical for those days?
- Do you get a "bounce-back" on days after public holidays?
- Are there monthly or seasonal effects?
- How does the volume vary at different times of the year, by month or by quarter?

One can use multiple regression methods to determine these effects and patterns. For example, one might start with our "typical" levels for a day. We then see that Mondays receive more calls than that, while Fridays receive fewer calls. Then, during school holidays, call volume might decrease further, and so on.

From this, one can start to predict how many calls we will get on any day in the future – for example on Friday, 12 August 2016. It will be known that it is a Friday in the middle of the month of August, in the third quarter of the year (Q3), during the school holidays, and it will be known how the number of calls per day in August, in Q3 etc., differs from that of a typical day – so therein lies the embryo of a forecast.

It might also be that, after allowing for these effects, there is a general increase over time, as a trend of daily calls – maybe for every month that elapses, you will get an additional number of calls per day. If this is observed, it might be sensible to understand what the reason for this is before applying it to future predictions. Is this general increase something that will continue in this way, or might it level out or reverse?

Then as one gets closer to the days being predicted (i.e. 4–6 weeks out), one can create shorter-term forecasts by using information on the numbers of calls received on recent days to enhance the prediction. An unexpectedly busy two weeks, say, might raise the chance of people needing to be on hold for longer. This might then have a knock-on effect on the number of calls coming in over the next few weeks.

From regression to forecasting

Table 1 shows the statistically significant "effects" obtainable from the regression models on historical days; note the linear decreasing trend effect ("week") and the fact that one level of each variable has been set as the "implicit zero effect" (e.g. Saturday and January). One can then take these effects and see whether or not they apply on any future day.

All days pick up the 4523 starting number of calls, which in this example is the typical number of calls on a Saturday in January in the first seven days of the month. Note the negative coefficient of 5.1, denoting an overall trend of each passing week receiving

Table 1. Statistically significant effects obtainable from regression models on historical days

Driver	*Meaning*	*Effect*
Intercept	Start point of calculation	4523
dow_2	Monday	1889
dow_3	Tuesday	709
dow_4	Wednesday	524
dow_5	Thursday	256
dow_6	Friday	107
dow_7	Saturday	0
dim_q1_7	1st–7th of month	0
dim_q8_14	8th–14th of month	–34
dim_q15_21	15th–21st of month	–93
dim_q22_28	21st–28th of month	–250
dim_q29_31	29th–31st of month	–214
mon_1	January	0
mon_2	February	–863
mon_3	March	457
mon_4	April	251
mon_…	…	…
mon_8	August	–280
mon_9	September	185
mon_…	…	…
mon_12	December	–1159
qtr_1	Quarter 1	0
qtr_2	Quarter 2	–175
qtr_3	Quarter 3	–341
qtr_4	Quarter 4	–570
SH	School holidays	–504
TABH	Tuesday after bank holiday	1500
week	For each week number	–5.071

5 fewer calls per day. Admittedly, this is a small effect in the short term, but it did come out as significant and therefore contributed positively to the overall predictive power of the model and is indicative of a fall of about 250 daily calls year-on-year. This trend gradient held true throughout the range of the data collection period, although it is good practice to determine whether different parts of the series show different growth/decline patterns when carrying out pre-modelling observations on the data and, if so, to test accordingly.

With our first example (Friday, 12 August 2016), one would subtract 341 calls for being in Q3 (a quiet time of year, generally) and a further 504 for being in the school holidays, along with the addition of 107 calls for being a Friday, and –34 on account of it being on the 12th of a month and –280 for August. An incrementing "week number" reference was given to each week, and the one covering that day is 241. For each number of the week, one would subtract 5.071 calls – in this case, that amounts to 5.071 × 241 = 1222. Therefore, overall, one is to expect 2249 calls on that day.

Six weeks later (on 19 September 2016), the schools are back and therefore the school holidays (SH) driver no longer applies, although the Q3 subtraction still applies. One would also apply the Monday flag (+1889), the 15th–21st of the month (–93) and the September flag (+185), giving a total of 4910 calls, once the week effect has been accounted for (5.071 × 247 = 1253).

The modelling methodology

The work described in this article relies on an autoregressive model, involving an intercept, a trend factor to allow for underlying increases and decreases, and various seasonal (month-in-year) and cyclical patterns of known frequency (e.g. weekly and monthly). All of the seasonal/date-based predictors are dummy variables (e.g. "month_1" for January, "month_2" for February, etc.).

The form of the model is:

$$Y_t = a + b_1x_{1t} + b_2x_{2t} + b_3x_{3t} + \ldots + b_px_{pt} + \alpha_0 + \alpha_1Y_{t-1} + \alpha_2Y_{t-2} + \ldots + \alpha_pY_{t-p} + e_t$$

Note that the b terms relate to the seasonal/day-of-the-week effects, while the α terms are the autoregressive elements.[2] Such an autoregressive model specifies that the output variable depends linearly on its own previous values. These are "brought into play" 4–6 weeks before the day being forecast.

This means that the call volume for 12 August 2016 could be predicted by accounting for the trend and seasonal factors exemplified above, along with observed call volumes on 11, 10, 9, … August. In practice, when forecasting for the 12th, the volume on the 11th (and closely preceding days) would not be known either, so instead, the predicted volume for the 11th is used as a proxy. Consequently, the further into the future one is running a short-term forecasting model, the greater the "proxy versus actual" component of the forecast and hence the greater the volatility.

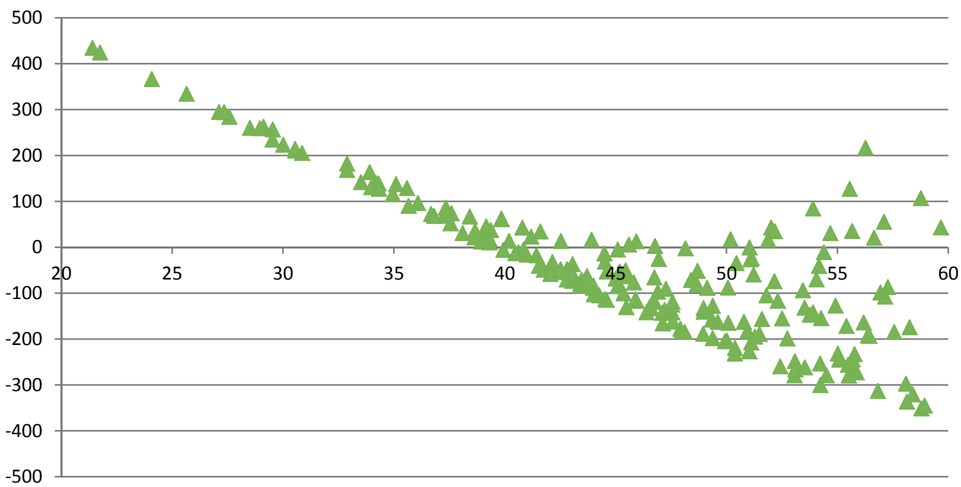

Figure 1. Model validation – unacceptable pattern in the residuals

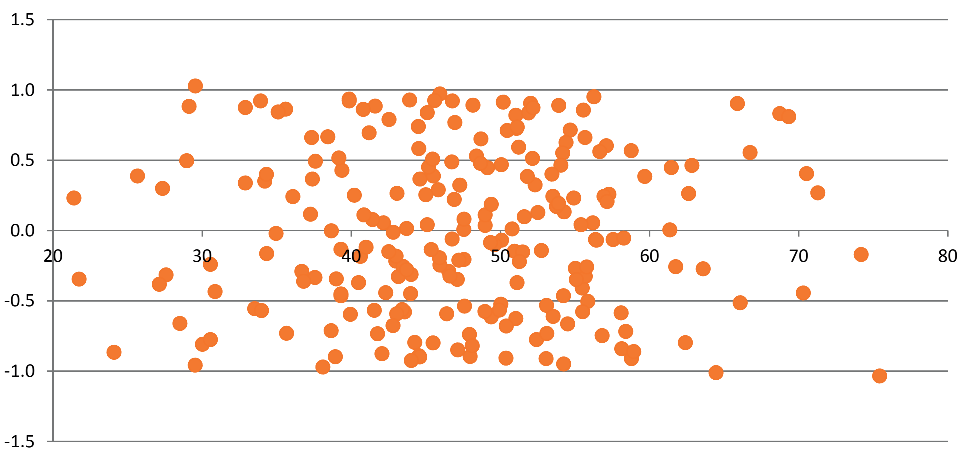

Figure 2. Model validation – acceptable pattern in the residuals

How accurate is the model?

Only time will tell how good your model is, and the only way of completely validating your forecasts is to wait until the day in question. However, if the model is suboptimal, then this may be too late as either the customer experience or the staffing budget may be on course to suffer.

However, there are a number of essential and familiar model validation procedures that should be carried out, such as reviewing the coefficient of determination (R^2), mean squared errors and Akaike's information criterion to assess the "goodness of fit" and "backtesting", which is widely used in financial applications (see bit.ly/1IiFamE). Needless to say that recognised (stepwise) model-selection procedures should have been carried out at the model construction stage so as to avoid over-fitting and variance inflation of the model. One should also look at the residuals (the differences between the historical values observed and what would have been predicted by the model) to verify that their variance is constant over the range of data upon which modelling has been conducted. Should the last of these be breached, the residuals would tend to "fan out" when plotted against the x-variable, as in Figure 1. An example of the pattern of residuals from a more acceptable model is shown in Figure 2. The residuals have zero mean throughout the range of the data.

Once you are happy with the model, the typical "envelope" of accuracy can be estimated by calculating the percentage variance of each historical month between the actual data and what your forecast would have been (see Table 2). The percentage variance is calculated by subtracting the forecast amount from the actual observation and dividing by the forecast. From this, one can work out the mean absolute variance (MAV) across all time points so as to see what proportion of MAVs are within the required accuracy range (say, ±5%).

The reason why the word "envelope" has been used is that an assessment of the accuracy of the model on the basis of how it would have performed on historical data is likely to be optimistic, since one is testing the model on the same data as was used to create the model, and this is referred to as "backtesting" (see above).[1]

A more robust approach would involve splitting the data randomly into two halves, one to create the parameters and the other to calculate the retrospective variances. However, this may be tricky in practice with small-base single-observation time series data and one might be pragmatic and settle for these variances simply indicating a "best case" scenario (i.e. how good your model could potentially be), hence the usage of the word "envelope" here.

Further refinements

By now, your model is likely to be improving, but there are some further effects that may be explored to fine-tune it. If, for example, there has been a steadily increasing trend in call volume, is this due to an increasing number of people that are eligible

Table 2. Comparisons of model predictions against actual data

		Accuracy envelope	MAV	2.1%
			% over 5%	9.1%
	A	P	$V=(A-P)/P$	Abs(V)
Date	Actual	Prediction	% variance	Abs variance
15Jan2015	4824	4662	3.5%	3.5%
16Jan2015	4439	4470	-0.7%	0.7%
17Jan2015	658	690	-4.6%	4.6%
18Jan2015				
19Jan2015	6120	6196	-1.2%	1.2%
20Jan2015	4942	5097	-3.0%	3.0%
21Jan2015	4705	4834	-2.7%	2.7%
22Jan2015	4471	4585	-2.5%	2.5%
23Jan2015	4405	4377	0.6%	0.6%
24Jan2015	694	687	1.0%	1.0%
25Jan2015				
26Jan2015	6312	6162	2.4%	2.4%
27Jan2015	5122	5035	1.7%	1.7%
28Jan2015	4852	4774	1.6%	1.6%
29Jan2015	4564	4643	-1.7%	1.7%
30Jan2015	4136	4428	-6.6%	6.6%
31Jan2015	701	683	2.6%	2.6%
01Feb2015				
02Feb2015	6329	6092	3.9%	3.9%
03Feb2015	4985	4963	0.4%	0.4%
04Feb2015	4976	4699	5.9%	5.9%
05Feb2015	4466	4508	-0.9%	0.9%
06Feb2015	4173	4295	-2.8%	2.8%
07Feb2015	701	693	1.1%	1.1%
08Feb2015				
09Feb2015	6073	6032	0.7%	0.7%
10Feb2015	4862	4906	-0.9%	0.9%
11Feb2015	4783	4638	3.1%	3.1%
12Feb2015	4497	4443	1.2%	1.2%
13Feb2015	4211	4233	-0.5%	0.5%
14Feb2015	615	695	-11.5%	11.5%
15Feb2015				
16Feb2015	5689	5653	0.6%	0.6%
17Feb2015	4413	4536	-2.7%	2.7%
18Feb2015	4265	4280	-0.3%	0.3%
19Feb2015	4206	4085	3.0%	3.0%
20Feb2015	3924	3869	1.4%	1.4%
21Feb2015	684	692	-1.1%	1.1%
22Feb2015				
23Feb2015	???	5852	???	???
24Feb2015	???	4725	???	???
25Feb2015	???	4464	???	???
26Feb2015	???	4272	???	???
27Feb2015	???	4054	???	???
28Feb2015	???	692	???	???
01Mar2015				
02Mar2015	???	6212	???	???
03Mar2015	???	5088	???	???
04Mar2015	???	4828	???	???
05Mar2015	???	4636	???	???
06Mar2015	???	4421	???	???
07Mar2015	???	690	???	???
08Mar2015				
09Mar2015	???	6174	???	???
10Mar2015	???	5042	???	???
11Mar2015	???	4773	???	???
12Mar2015	???	4581	???	???
13Mar2015	???	4368	???	???
14Mar2015	???	691	???	???

to call in, or an increase in the likelihood of a given person calling in, or both? With accurate forecasts of changes in the population base – in this case, the number of people insured by the private healthcare firm – could the forecasts be improved by modelling the number of calls per 1000 people per day, then multiplying this by the eligible population forecast?

On a related note, we can split the forecast into two further components – each behaving in different, yet predictable ways (bit.ly/1SBuBfL): the number of individual "unique" callers calling in on a given day, and the ratio of all calls to unique calls – otherwise known as the "spin factor".

The number of unique calls is driven fairly predictably by seasonal and day-of-week factors and forms the mainstay of long-term models, while the spin factor will depend heavily on the number of unanswered calls from recent days. Indeed, many call centres get periods where demand cannot be met, and this leads to a downward spiral of increased call abandonment and additional (unmet) demand, despite the fact that the number of unique callers will have remained relatively unchanged.

During stable periods, 4000 individuals may generate 4300 calls in a day, but during the peaks, those 4000 callers may generate as many as 4800 calls. Thus, by forecasting in this way, one can obtain an accurate idea of the size, duration and severity of the call peak and determine how much emergency resource to invoke and for how long.

Customers do not like waiting on the phone, but statistical models can go some way towards helping call centre managers tailor staffing levels to expected demand

There is also the question of how the spikes or step-changes might be dealt with in the historical time series. One would first naturally need to have carried out appropriate techniques to screen for outliers in the data (1.usa.gov/1mneyrN), especially atypically high or low observations adjacent to more expected ones, and would need to have understood the reasons underlying these changes (e.g. faulty measurement of volume, temporary closure of the service).

If there was a sudden 30% increase in call volume 6 months ago as the operation took over an additional area of service, it would be tempting to play safe and only use the comparable past 6 months to develop the forecasts. However, by building in, testing the significance of and deriving a parameter for a dummy variable at all points after this takeover, there exists a sound alternative to discarding earlier information that would otherwise be highly informative in showing critical drivers of call volume, such as seasonal and day-of-week, which are needed to create an accurate forecast.

If one is anticipating a marketing activity that will stimulate call demand, then there may have been similar situations within the

- Variance between actual and predicted workload/volumes i.e. (Actual – Forecast / Forecast) as a %;
- % of calls not abandoned;
- % of calls handled within X seconds?
- Average waiting times;
- Client satisfaction levels;
- Cost of running service;
- Efficiency of running service;
- Staff satisfaction/engagement levels;
- Customer retention, revenue and profit (private sector).

Basic measures

Effect on organisation

Figure 3. Measuring call forecasting performance

history. By assessing how volumes behaved when these conditions applied and expressing this as an upward or downward effect (e.g. 250 extra calls per day), it would be possible to apply such changes at points in time when a similar situation is likely to occur again.

Practical considerations

The more successful models are those where one has started the process by thinking about all of the factors that might potentially influence call volumes, rather than by what data appears easily available. Having generated an overall list, one should consider where the information may be obtained, how it might be collected, and what the barriers are for using it. This may appear particularly challenging for developing new services, although the principles are much the same. Surveys of potential users may play a useful part, along with reviewing what has happened with similar services. If the demographic make-up of an area with and without an existing service is the same, then the calls per 1000 members per day figure could be lifted across.

Measures of performance of your forecasts are numerous, and as with many other key performance indicators, range from the purely process-driven (such as the variances between the workload forecast and what actually occurred) to those that are far more fundamental to determining the actual impact that the forecasts are having on the overall performance of the business or service. These are listed in Figure 3. Are calls being answered in time? Are callers happy with the service and is it all within budget?

Customers do not like waiting on the phone, but statistical models can go some way towards helping call centre managers tailor staffing levels to expected demand. Of course, some planners may wish to continue using their "gut feel" to determine call volumes and staffing levels. This can yield accurate results intuitively. However, in most cases, it should still be possible to capture and document these "gut feel" processes and to bring these factors into a statistical model in order to obtain the best of both worlds.

Using the pragmatic techniques described in this article, it should theoretically be possible to forecast most things with a degree of accuracy that can exceed those of straight averages or best guesses.

How far back and how far forward?

The further back the historical time series goes, the greater the level of confidence in the overall trend, the seasonal and other repeating effects, and hence the greater the forecast precision.

For seasonal effects, a minimum of three years would be needed in order to inform whether, for example, Januarys are generally busier than other months.[3] Three cycles would at least provide some measure of variation/precision of the seasonal estimates. If less than a year is available, then daily forecasts (on daily histories) may still be possible, using trend and day-of-the-week effects.

Long-term forecasts based purely on seasonality and historical trends can be projected several months into the future. However, the inherent assumption that the strength and directionality of these effects will still hold true in the future, as they have in the past, will need to be revalidated. Thus the model parameters should be recalculated on a regular basis – ideally every few months.

However, with short-term autoregressive forecasts – where the estimates of volumes on day X are dependent on those on previous days – model assumptions may only hold good for a month into the future.

In practice, long-term forecasts are created initially. Then, as one gets closer to the day in question and where autoregressive models are acceptable (i.e. up to 4–6 weeks ahead), then the longer-term models are overwritten by shorter-term ones, which take account of a greater range of factors and should be more precise.

References

1. James, K. E., White, R. F. and Kraemer, H. C. (2005) Repeated split sample validation to assess logistic regression and recursive partitioning: an application to the prediction of cognitive impairment. *Statistics in Medicine*, **24**, 3019–3035.
2. Pickett, J. C., Reilly, D. P. and McIntyre, R. M. (2005) How to select a most efficient OLS model for a time series data. *Journal of Business Forecasting*, **24**(1), 11–15. Available at bit.ly/1HISU8H
3. Webby, R. and O'Connor, M. (1996) Judgemental and statistical time series forecasting: a review of the literature. *International Journal of Forecasting*, **12**(1), 91–118.

Andrew Zelin is a senior analytics manager at BUPA, with 20 years' experience as a professional statistician, running analytical projects in market research, central government, health and the telecoms sectors. Alongside this, he gives training for the Market Research Society and other training companies. He has presented and published papers in a range of statistical and survey-related practices, at Royal Statistical Society and AAPOR conferences and published two award-winning methodology papers. He can be reached by email at: andrew3193@live.co.uk