WHO IS DUE BACK? PART IV:

APPLYING BI TO MEASURE MARKETING RESULTS

By Dr. A. K. Singh and Andrew Cardno

Authors' Note: So far in this series, we have looked at how due-back can be applied to measure and predict customer behavior. It was originally supposed to be a three-part series, but after receiving several questions from readers and comments on a CEM blog entry relating to how to measure effectiveness of marketing programs (www.casinoenterprisemanagement .com/blog/theoretical-win-taking-closer-look-sacred-cow), we decided to expand the series to five parts.

The world of marketing is full of rules of thumb. For example, according to *Web Marketing for Dummies*, "As a rule of thumb, spend no more than 10 percent of your average sales amount on advertising." While these "rules" can be powerful mechanisms for making "practical" decisions, we're going to dig deeper and show how we can apply modeling instead to find an optimal amount of marketing effort. Further, we will show how you can use business analytics to determine if marketing promotions significantly impact win in a positive way.

Multiple Regression

Multiple regression is one of the basic tools of statisticians. It provides a mechanism whereby a function is built that can either interpolate or extrapolate, but this function is fitted in such a way as to minimize the sum of the squares of the errors from the model.

A positive regression coefficient means that that the predictor increases the value of the outcome, while a negative regression coefficient means that it decreases the value of that outcome. A large-in-magnitude regression coefficient means that the predictor strongly influences the outcome, while a near-zero regression coefficient means that the predictor has little influence on the value of the outcome.

In multiple regression, the "total sum of squares" of the response (output value) is split into two parts: 1) "sum of squares due to regression," which represents the effect of the fitted model or "signal"; and 2) the "sum of squares due to error," which represents the "noise." The term R^2, which relates to the correlation coefficient, is essentially a calculation based on these two sums of the squares.

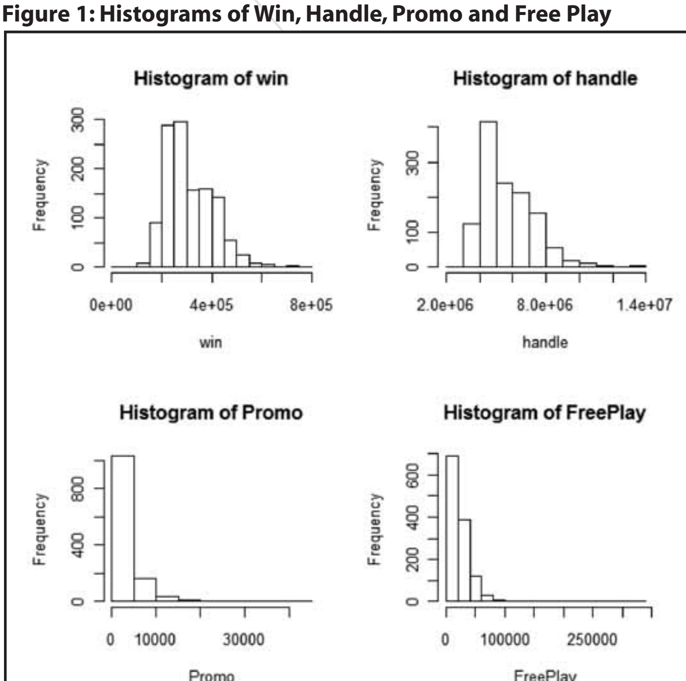
The Data

The data we used for this illustrative example is from a small casino in Las Vegas. A subset of the data is shown in Table 1. Figure 1 shows histograms of win, handle, promotion and free play amounts.

Table 1: A Subset of Time Series of Handle and Win Amounts

Date	Day	Free Play	Handle	Win	Promo
9/1/2009	Tuesday	9461	3797369	162378	0
8/31/2009	Monday	37709	3401280	212064	0
8/30/2009	Sunday	49258	4505782	265852	0
8/29/2009	Saturday	36500	6174634	403711	7313
8/28/2009	Friday	65136	6342854	448521	2698
. .	.	.	.	.	
. .	.	.	.	.	
. .	.	.	.	.	
8/1/2009	Saturday	58745	5944857	387395	6674
7/27/2009	Monday	46900	4650129	381169	0
7/26/2009	Sunday	37774	5399841	115231	11131
7/25/2009	Saturday	5440	6395415	423838	6603

Figure 1: Histograms of Win, Handle, Promo and Free Play



Tables 2 and 3 show the mean and median of win by day of week and promotion (no, low, medium and high). Both mean and median win for the no promo and high promo categories are very similar, which seems to suggest that marketing promotions are not working for this casino.

It can also be seen from Tables 2 and 3 that the mean and median win are high during the weekends across the four promo categories, which is expected.

Figure 2 shows a time series graph of the data. As can be seen from the graph, the data it is quite variable. The determination of the effects of various sources of variability in the data requires a smorgasbord of statistical tools.

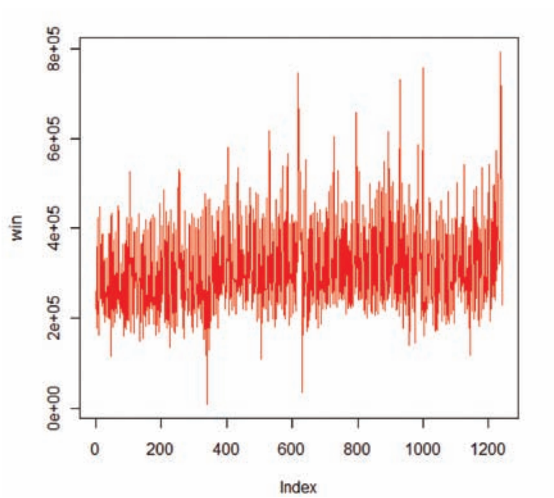
Table 2: Mean Win by Day of Week and Level of Promotion

Day	No Promo	Low Promo	Med. Promo	High Promo
Monday	255146	232734	257543	246483
Tuesday	242197	220761	244754	256602
Wednesday	262858	239264	257828	273607
Thursday	266223	226183	260113	272709
Friday	402122	371231	397635	416387
Saturday	441823	412282	425675	421806
Sunday	327627	253733	325553	316450

Table 3: Median Win by Day of Week and Level of Promotion

Day	No Promo	Low Promo	Med. Promo	High Promo
Monday	243397	222573	240665	246546
Tuesday	239655	217869	242566	255642
Wednesday	255807	228966	252685	261504
Thursday	252605	215551	257615	264917
Friday	397915	390036	396574	415092
Saturday	437754	419004	408974	419758
Sunday	314759	271262	308556	317353

Figure 2: Time Series Plot of Win



The Results

Multiple linear regression was used to fit a model to win as a function of the predictors (levels of promotion, free play amount, day of week, and time variables t and t^2 or t-square). The results are shown in Table 4. R-square value for the fitted model (58 percent) is not high, but it is still acceptable. Quite simply, this means that the regression model can determine the revenue quite accurately before it happens. It can also measure the relative strength of various marketing programs.

Table 4 shows that for the variable promo, only high promo has a P-value of .03, and hence, the high promo term is statistically significant, at 5 percent error. All other promo terms (DP1, DP2) have P-values exceeding .05 and, hence, are not significant. Negative coefficients for Monday through Thursday, and positive coefficients for Friday and Saturday (Sunday was used as the baseline), suggest that win tends to be higher over the weekends and lower during the weekdays. The time variables t and t^2 are also significant, indicating the presence of a quadratic trend in win values over time.

Actionable Insights

- Medium-sized promotions have as much impact as Fridays on business.
- Large promotions have more impact than Saturdays on business.
- Small promotions have a negative effect on business.

Now, using this model we can then take the next step and determine the effectiveness of a particular marketing program, as compared to other programs of a similar size. For example, we may run a small promotion and see that it is as effective as a medium-sized promotion.

The residuals from the model are auto-correlated, but we did not include the temporal terms for this article. The next session discusses the background to the temporal model.

Table 4: Multiple Linear Regression Model for Win

Predictor Variable		Estimate	SE	t	P-value
Zero Promotion	(Intercept)	309000	6763	45.691	0
Low Promo	DP1	-14510	7798	-1.86	0.0631
Med. Promo	DP2	8180	5499	1.488	0.1371
High Promo	DP3	11200	5168	2.167	0.0304
Free Play	Free Play	1	0.12	5.341	0
Monday	Dmon	-67860	6832	-9.933	0
Tuesday	Dtue	-72600	6886	-10.542	0
Wednesday	Dwed	-56250	6854	-8.206	0
Thursday	Dthur	-52150	6858	-7.605	0
Friday	Dfri	72440	7094	10.212	0
Saturday	Dsat	102200	6921	14.762	0
Time (days)	t	19640	2125	9.245	0
Time (days)-Squared	t^2	-5892	2323	-2.536	0.0113

Removing Seasonality with ARIMA

One of the great challenges with time series data is taking into consideration sources of variation, such as seasonality and random shocks. While we can use rules of thumb like comparisons to the same time last year, statistical models can take this much further. While it might sound esoteric, Auto-Regressive Integrated Moving Average (ARIMA) modeling is extremely useful for this.

The Still Wheel Effect

To explain ARIMA analysis, imagine a wheel on a car. If you were interested in determining whether the car was going up or down a hill, and the only way you could do this way by placing a dot on one of the wheels and tracking the movement of that dot, then you could find your answer by measuring to see if the dot was rising or falling. But the problem is that as the wheel spins, the dot moves up and down, and this "noise" could be greater than the movement up or down the hill. To properly analyze the movement of the car, we need to remove the effect of the wheel's spin.

If the wheel spins 30 times per second and your camera is taking 30 frames per second, the wheel looks like it is not moving—except for the dot! Now you can simply see if the point is moving up or down. *[Note: You can see the "still wheel effect" in action in lots of old movies.]*

Customers are Wheels Within Wheels

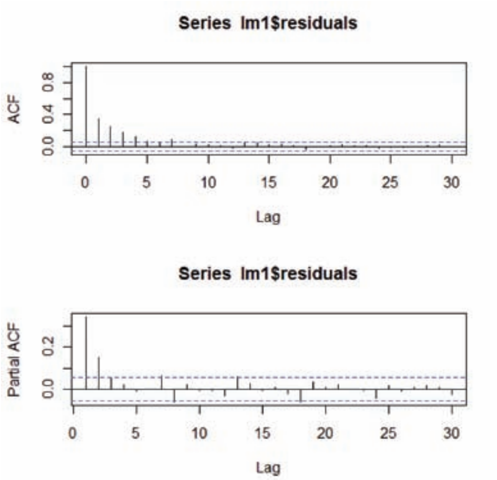
Measuring trends in revenue is like measuring wheels within wheels. The smallest wheels are daily cycles, and the larger wheels are weekly cycles. Larger yet wheels are yearly cycles, and maybe the huge wheels are seven-year cycles. If we can remove all the effects of these cycles, then we can say with more accuracy whether the business is trending up or down. We can also determine if the observed effects of a marketing program are from the actual program or are likely to really be the result of a cyclical effect.

The First Steps

In Figure 3 we present two graphs. The first is the autocorrelation function between a lag period (measured in days) and the current date. Looking at the graph, we can say that the time series data is not correlated after we look back more than five days.

The second graph in Figure 3 is the partial autocorrelation function[1], which is an analysis of the autocorrelation by not considering days less that the lag period (again in days). The ACF and PACF by inspection give an indication of the number of terms required in the ARIMA model.

Figure 3: Autocorrelation and Partial Autocorrelation Functions of the Residual from Fitted Model



The Goals

The first goal of temporal analysis is to discover the impacts of other factors, such as the effect of a marketing campaign. The second goal of temporal analysis is to discover the nature of the temporal patterns; for example, if we can see that the data exhibits a mainly weekly pattern, then that might push us toward using more of a weekly-based marketing program (many programs run on a monthly cycle).

Finding the Exceptions

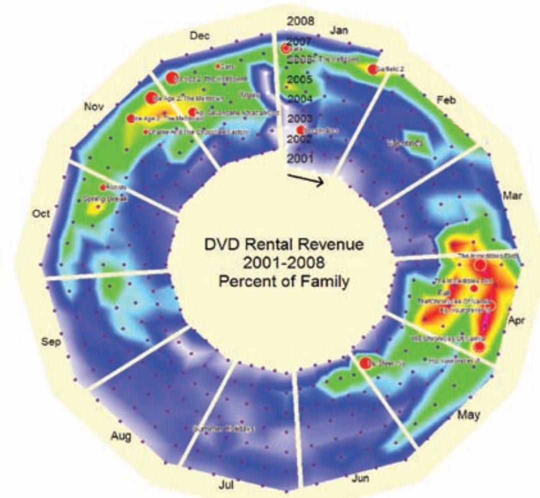
While finding the "stillness in the wheel" using computerized ARIMA models and lag analysis, there is still a role for humans in this process. The role of the human is to provide insight to see the exceptions in the data and to be able to understand and communicate the different patterns in an actionable way across the business—without having to learn the complex art of ARIMA modeling.

Figure 4 shows a wheel-shaped calendar comparing the percentage of revenue of a DVD rental business for the "family" category of DVDs from 2001 to 2008. Red colors are hot (or high), blues are cold (or low).

By comparing the center of the circle with the weekly dots as you move your eye out of the circle, you can see the "stillness of the wheel."

As is shown in Figure 4, we can present the data on the wheel; this, in some cases, might bypass the need for ARIMA modeling. (For example, in this model the yearly hotspot patterns are quite apparent to the observer. This wheel-shaped graphic shows a "still wheel effect" near April, November and December.)

Figure 4: Cyclical Data



Final Thoughts

When business is increasing and the market is growing, it is easy to claim success for the effect of any marketing program, as all the programs seem to "work." But, in fact, the success of the business may be due to some other factor, not the marketing. However, when the market is not growing, measurement of success is even more difficult, as all programs seem to "fail" in this environment. This is when using more sophisticated models and visualization techniques is critical in determining the net effect of the marketing programs and in finding new opportunities.

While it is a tough environment, surely it will be those with the strongest vision and the best judgment on how to use the correct BI tools who will be the winners … and maybe a few lucky ones, too. The question of the day is would you rather rely on luck or business intelligence?

1 www.duke.edu/~rnau/411arim.htm.

ANDREW **CARDNO**

Andrew Cardno has more than 16 years of experience in business analytics, ranging from modeling health care drive times to casino gaming floor analytics. He often presents on the future of analytics across the world and has spent the last seven years living in the United States and working with corporations around the world. He can be reached at andrewcardno@yahoo.com.

DR. A. K. **SINGH**

Dr. A. K. Singh has taught statistics, mathematics and operations research courses at New Mexico Tech, Socorro, N.M., and statistics and mathematics courses at University of Nevada, Las Vegas. He has more than 75 publications in theoretical and applied statistics and can be reached at aksingh@unlv.nevada.edu.