# Forecasting Casino Gaming Traffic with a Data Mining Alternative to Croston's Method

*Barry King*

**Abstract**

Other researchers have used Croston's method to forecast traffic at casino game tables. Our data mining alternative to Croston's method more accurately forecasts gaming traffic using the rich databases that are frequently available at casinos. A more accurate forecast will allow for better planning of staffing on the casino floor.

*Keywords*: Croston's method; gradient boosting method; R; random forest

## Introduction

Chen, Tsai, and McCain (2012) report research where demand for a game table at a minimum betting limit is forecast using Croston's method for intermittent demand. The forecasted demand is allocated to game tables using a mixed integer linear programming (MILP) model to maximize theoretical hold. This research presents an alternative data mining approach for forecasting casino table demand that is more accurate than Croston's method and that uses multiple covariates in a data mining forecast. The MILP model is not discussed.

This paper is structured as follows: 1) a review of some of the literature on forecasting intermittent demand; 2) description of the hypothetical casino that is referred to throughout the paper; 3) description of the forecasting problem; 4) a detailed discussion of the simulated-demand engine; 4) examination of Croston's method followed by the two-stage data mining forecasting approach; and 5) presentation of results ending with concluding remarks.

## Literature Review

Croston (1972), as augmented by Rao (1973), is seminal to the research on forecasting intermittent demand. Syntetos and Boylan (2005) present a correction to Croston where in it they observe that Croston's method is biased, and that the magnitude of the error depends on the smoothing constant value being used. McCabe and Martin (2005) discuss a Bayesian approach to addressing low count time series. Callagaro (2010) is an easily readable and thorough examination of methods for forecasting spare parts demand. The paper reports that Croston's method has a smaller positive bias when few demands are zero and that the Syntetos and Boylan modification has a smaller bias when many demands are zero. Snyder, Ord, and Beaumont (2012) discuss modelling approaches to the slow moving inventories problem. Numerous methods are discussed. They observe that the Syntetos and Boylan correction to Croston's method provides superior point forecasts for "faster intermittent" items, those with relatively short mean times between non-zero demand periods. Kandananoud (2012) compares various forecasting methods for autocorrelated time series.

Barry King
Associate Professor
Andre B. Lacy School of Business
Butler University
email: king@butler.edu

This research offers an accessible, data mining alternative to Croston's method for forecasting demand at casino table games. It is especially appropriate as an alternative forecasting method for the Chen, Tsai, and McCain (2012) article.

**Example Casino**

In order to examine the two techniques for forecasting traffic at casino game tables, Croston's method and data mining, a rich data set is required. From proprietary casino data of an American casino, we acquired characteristics for two games: mini-baccarat and poker. Table 1 and Table 2 report the results of this acquisition.

Table 1
*Partial Description of a Casino Floor*

| Asset Number | Pit | Table Number | Capacity | Hands per Hour | Theoretical Hold | Game Code | Game Description |
|---|---|---|---|---|---|---|---|
| 10001861 | 4 | 1 | 6 | 72 | 0.011 | MIN-BAC | Mini Baccarat |
| 10002542 | 8 | 4 | 6 | 72 | 0.011 | MIN-BAC | Mini Baccarat |
| 10002546 | 8 | 5 | 6 | 72 | 0.011 | MIN-BAC | Mini Baccarat |
| 10002475 | 4 | 12 | 6 | 72 | 0.011 | MIN-BAC | Mini Baccarat |
| 10002981 | 5 | 1 | 10 | 23 | 0.034 | POKER | Poker |
| 10002942 | 10 | 1 | 10 | 23 | 0.034 | POKER | Poker |
| 10002982 | 5 | 2 | 10 | 23 | 0.034 | POKER | Poker |
| 10002948 | 6 | 2 | 10 | 23 | 0.034 | POKER | Poker |
| 10002944 | 10 | 3 | 10 | 23 | 0.034 | POKER | Poker |
| 10002950 | 6 | 4 | 10 | 23 | 0.034 | POKER | Poker |
| 10002943 | 10 | 4 | 10 | 23 | 0.034 | POKER | Poker |
| 10002983 | 5 | 5 | 10 | 23 | 0.034 | POKER | Poker |
| 10002945 | 10 | 6 | 10 | 23 | 0.034 | POKER | Poker |

*Note*. Asset Number is the unique identification number of the game table. Pit and Table Number indicate where the Asset is located on the gaming floor. Theoretical Hold is the average house advantage.

Table 2
*Game and Minimum Bet Categories*

| Game Code | Minimum Bet (USD) | Game and Minimum Bet Category |
|---|---|---|
| MIN-BAC | 50 | BAC_0050 |
| MIN-BAC | 150 | BAC_0150 |
| MIN-BAC | 200 | BAC_0200 |
| MIN-BAC | 1000 | BAC_1000 |
| POKER | 2 | PKR_0002 |

*Note*: A player who sits at a $200 minimum bet mini-baccarat table must be willing to bet at least $200 at each hand. Although casino poker games can involve large sums of money, typical ante is $2.

Mini-baccarat is a six seat game that uses eight decks of cards and is based on American baccarat. It is strictly a game of chance, with no skill or strategy involved. A standard poker table has ten seats. A variety of games are played at a poker table.

**Forecasting Problem**

Casino patrons may arrive at casino game tables for a given minimum bet at a near random rate. The analyst's task is to 1) forecast if there will be demand at the table for a given minimum bet and then 2) if there is positive demand, forecast the demand for that hour (see Figure 1 and Figure 2). Headcount can be estimated by using loyalty cards; however, not every player has a loyalty card, and pit personnel may be too busy at times to gather the data so the count will show low. Alternatively, headcount also can be gathered by personnel in the security video booth.

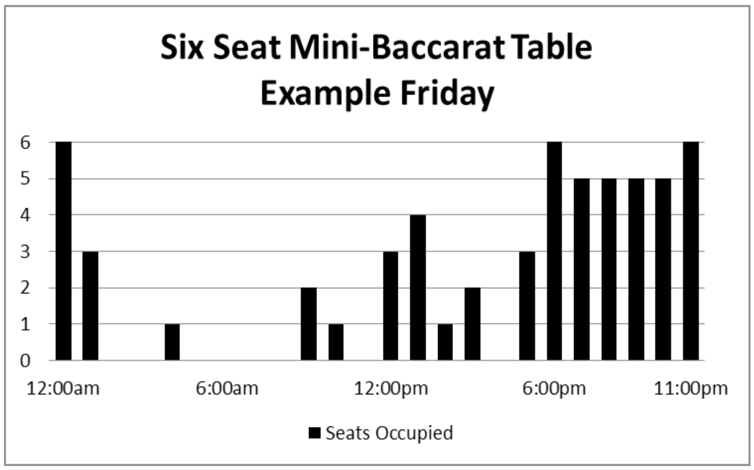


*Figure 1*. Example of intermittent demand and demand for a mini-baccarat table on a Friday.

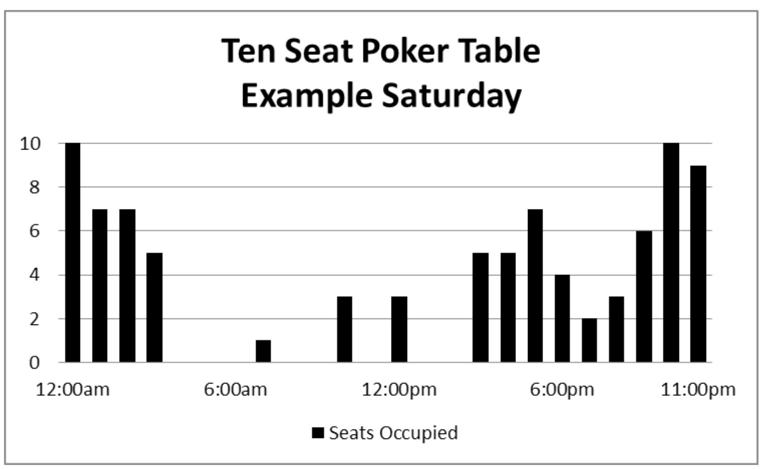


*Figure 2*. Example of intermittent demand for a poker table on a Saturday.

## Simulated Demand

One thousand weeks of simulated data were constructed for training data. Each week consisted of seven twenty-four hour days for each of two games, mini-baccarat and poker. A second 1000-week data set was constructed for testing.

This study does not distinguish between different betting limits per game code, unlike the model of Chen, Tsai, and McCain (2012). A link to the code and data for the demand simulator appears in Appendix A.

Three elements made up the simulation 1) expected traffic density by day and by hour (see Table 3), 2) probability distributions of the number of seats occupied by traffic density and by game table (see Table 4), and 3) an autocorrelated uniform probability random number generator in order to effect producing the number of seats occupied from the probability distributions table (see Figure 3). The random number generator needed to produce autocorrelated values in order to simulate demand streams where the number of patrons at a table at time t+1 was related to the number at the table at time t.

Table 3
*Expected Traffic Density*

| Hour | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
|---|---|---|---|---|---|---|---|
| 12:00 AM | TD-10 | TD-5 | TD-5 | TD-5 | TD-6 | TD-6 | TD-10 |
| 1:00 AM | TD-10 | TD-3 | TD-3 | TD-3 | TD-5 | TD-5 | TD-9 |
| 2:00 AM | TD-9 | TD-3 | TD-3 | TD-3 | TD-4 | TD-4 | TD-8 |
| 3:00 AM | TD-8 | TD-2 | TD-2 | TD-2 | TD-4 | TD-4 | TD-7 |
| 4:00 AM | TD-5 | TD-2 | TD-2 | TD-2 | TD-3 | TD-3 | TD-6 |
| 5:00 AM | TD-3 | TD-1 | TD-1 | TD-1 | TD-3 | TD-3 | TD-4 |
| 6:00 AM | TD-2 | TD-1 | TD-1 | TD-1 | TD-2 | TD-2 | TD-4 |
| 7:00 AM | TD-1 | TD-2 | TD-2 | TD-2 | TD-2 | TD-2 | TD-3 |
| 8:00 AM | TD-2 | TD-2 | TD-2 | TD-2 | TD-1 | TD-1 | TD-3 |
| 9:00 AM | TD-2 | TD-3 | TD-3 | TD-3 | TD-2 | TD-2 | TD-4 |
| 10:00 AM | TD-3 | TD-3 | TD-3 | TD-3 | TD-2 | TD-2 | TD-5 |
| 11:00 AM | TD-3 | TD-4 | TD-4 | TD-4 | TD-3 | TD-3 | TD-5 |
| 12:00 PM | TD-4 | TD-5 | TD-5 | TD-5 | TD-4 | TD-4 | TD-5 |
| 1:00 PM | TD-4 | TD-5 | TD-5 | TD-5 | TD-5 | TD-5 | TD-6 |
| 2:00 PM | TD-5 | TD-6 | TD-6 | TD-6 | TD-6 | TD-6 | TD-6 |
| 3:00 PM | TD-5 | TD-6 | TD-6 | TD-6 | TD-6 | TD-6 | TD-7 |
| 4:00 PM | TD-6 | TD-5 | TD-5 | TD-5 | TD-5 | TD-6 | TD-7 |
| 5:00 PM | TD-6 | TD-5 | TD-5 | TD-5 | TD-5 | TD-7 | TD-8 |
| 6:00 PM | TD-7 | TD-6 | TD-6 | TD-6 | TD-6 | TD-8 | TD-8 |
| 7:00 PM | TD-8 | TD-7 | TD-7 | TD-7 | TD-7 | TD-8 | TD-8 |
| 8:00 PM | TD-8 | TD-7 | TD-7 | TD-7 | TD-8 | TD-9 | TD-9 |
| 9:00 PM | TD-9 | TD-8 | TD-8 | TD-8 | TD-9 | TD-9 | TD-9 |
| 10:00 PM | TD-9 | TD-8 | TD-8 | TD-8 | TD-9 | TD-9 | TD-10 |
| 11:00 PM | TD-7 | TD-5 | TD-5 | TD-5 | TD-10 | TD-10 | TD-10 |

*Note*. Traffic density ranges from the lightest density, TD-1, to the heaviest density, TD-10.

Table 4
*Probability Distributions*

Six Seat Mini-Baccarat

| | TD-1 | TD-2 | TD-3 | TD-4 | TD-5 | TD-6 | TD-7 | TD-8 | TD-9 | TD-10 |
|---|---|---|---|---|---|---|---|---|---|---|
| P(X=x) | X | x | x | x | x | x | x | x | x | x |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 |
| 0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 4 |
| 0.3 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 5 |
| 0.4 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 5 | 5 |
| 0.5 | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 5 | 5 | 5 |
| 0.6 | 0 | 0 | 1 | 2 | 3 | 4 | 5 | 5 | 5 | 6 |
| 0.7 | 0 | 1 | 2 | 3 | 4 | 5 | 5 | 5 | 5 | 6 |
| 0.8 | 1 | 2 | 3 | 3 | 4 | 6 | 6 | 6 | 6 | 6 |
| 0.9 | 1 | 2 | 3 | 4 | 5 | 6 | 6 | 6 | 6 | 6 |
| 1 | 2 | 3 | 4 | 5 | 5 | 6 | 6 | 6 | 6 | 6 |

Ten Seat Poker

| | TD-1 | TD-2 | TD-3 | TD-4 | TD-5 | TD-6 | TD-7 | TD-8 | TD-9 | TD-10 |
|---|---|---|---|---|---|---|---|---|---|---|
| P(X=x) | X | x | x | x | x | x | x | x | x | x |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 |
| 0.1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 4 |
| 0.2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 4 | 5 |
| 0.3 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 4 | 5 | 6 |
| 0.4 | 0 | 0 | 0 | 0 | 2 | 3 | 4 | 5 | 6 | 7 |
| 0.5 | 0 | 0 | 0 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0.6 | 0 | 0 | 1 | 3 | 3 | 5 | 6 | 7 | 8 | 9 |
| 0.7 | 0 | 1 | 2 | 4 | 4 | 6 | 6 | 7 | 9 | 10 |
| 0.8 | 1 | 2 | 3 | 5 | 5 | 6 | 6 | 7 | 9 | 10 |
| 0.9 | 1 | 3 | 4 | 5 | 6 | 6 | 7 | 8 | 10 | 10 |
| 1 | 2 | 4 | 5 | 6 | 6 | 7 | 8 | 9 | 10 | 10 |

*Note*. For example, if the value of the autocorrelated uniform random variate rounded to one decimal digit is 0.7 then a demand for 4 seats will be generated for a ten seat poker table during an hour of the day exhibiting TD-5 traffic density.

**Autocorrelated Uniform Random Numbers**

Willemain and Desautels (1993) report a method to construct autocorrelated uniform random numbers at a user selected correlation. See Appendix B for a description of this method. The Willemain and Desautels article displays a graph relating parameter c to serial correlation. We chose c = 1.15 which corresponds to a serial correlation of about 0.6. Figure 3 shows the autocorrelation of the uniform time series used in this study.

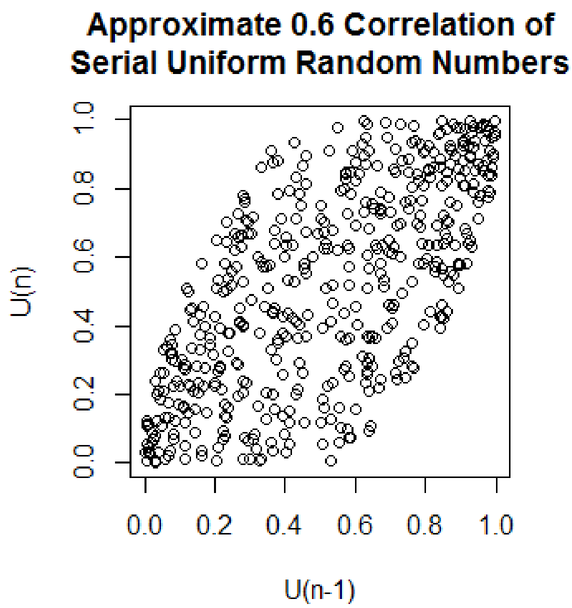


*Figure 3*: Autocorrelated uniform random numbers used to produce the simulated demand stream.

**Comparing Croston's Method to a Data Mining Method for Forecasting Casino Table Demand**

**Croston's Method**

Croston (1972) presents a method for forecasting intermittent demand in a supply chain environment. It forecasts the time between non-zero demand and then, for non-zero demand time slots, forecasts the size of the demand. Exponential smoothing is employed for these two tasks. It is necessary to specify a smoothing constant. The smoothing constant, α, was set to 0.30 for this study.

An assumption of Croston's approach is that the time between non-zero demands is intermittent, not random, and is therefore forecastable (see Figure 1 and Figure 2). Croston's method does not consider any covariates; it is strictly a time series analysis.

Definitions

- $X_t$ actual demand in period t
- $X'_t$ predicted demand at time t
- $T_t$ exponentially smoothed number of periods between non-zero demands
- $Z_t$ exponentially smoothed size of demand when it is non-zero
- $q$ number of periods since last non-zero demand
- $\alpha$ smoothing constant

Initialize

$Z_0 = X_0$

$T_0 = 0$

$q = 1$

Update (if zero demand)

$Z_t = Z_{t-1}$

$T_t = T_{t-1}$

$q = q + 1$

Update (if non-zero demand)

$Z_t = \alpha(X_t - Z_{t-1}) + (1 - \alpha)Z_{t-1}$

$T_t = \alpha(q - T_{t-1}) + (1 - \alpha)T_{t-1}$

$q = 1$

Forecast (if $T_t = 0$)

$X'_t = 0$

Forecast (if $T_t > 0$)

$X'_t = Z_t / T_t$

**Data Mining Method**

Contemporary casinos capture a great deal of data from the gaming floor, usually by patron who is using a loyalty card of some denomination. Since there is more information to support a forecast than just the time series of demand, data mining methods that use these data can be employed.

Today's data mining toolbox has many tools to assist analysts in understanding the data with which they are working. Tools such as neural networks, decision trees, support vector machines and many others are available. This study uses two ensemble methods: gradient boosting and random forests. The goal of ensemble methods is to combine the predictions of several base estimators built with a given learning algorithm in order to improve generalizability / robustness over a single estimator (Scikit-learn, n.d.).

A gradient boosting model was constructed to classify if a time slot will exhibit zero demand or non-zero demand. Gradient boosting is appropriate for binary classification and was chosen for this reason. A gradient boosting method is a machine learning decision tree. In boosting methods, base estimators are built sequentially and one tries to reduce the bias of the combined estimator. The motivation is to combine several weak models to produce a powerful ensemble (Scikit-learn, n.d.). The method is robust, that is it performs well even if its assumptions (i.e. the type of effects that are to be used and uniform distribution) are somewhat violated by the true model from which the data were generated (Decision tree advantages, n.d.).

A random forest regression was constructed for predicting the size of non-zero demand. In averaging methods, such as random forest, the driving principle is to build several estimators independently and then to average their predictions. On average, the combined estimator is usually better than any of the single base estimators because its variance is reduced (Scikit-learn, n.d.). Random forest was selected as the regression technique since it can work with a large number of predictor variables without first deselecting variables as well as not requiring cross validation (Breiman and Cutler, n.d.) and random forests tend not to overfit the data (Liaw & Wiener, 2002).

## Results

### Croston's Method

See the Value for Croston Approach column of Table 5. This column reports result statistics for applying the Croston method to the training and testing data sets. A Misclassification Rate of Empty or Occupied Seats is meaningless (NA) for the Croston approach. A positive mean error of 0.54 seats is not unexpected. Syntetos and Boylan (2005) demonstrate and correct for this mean error when there is a large number of zero-demand hours.

Of note are the rows for the Statistics for Seats Occupied ± One Seat area. This area describes statistics for predictions that are one seat off from the actual. For example, if the actual number of seats occupied for an hour is 7, then any forecasted value within one seat, 6, 7, or 8, will be counted as an accurate forecast since a forecast being off by just one seat is an acceptable tolerance.

Table 5
*Croston's Method*

| Statistic | Value for Croston Approach | Value for Data Mining Approach |
|---|---|---|
| Misclassification Rate of Empty or Occupied Seats | NA | 0.11 |
| Statistics for Seats Occupied | | |
| Mean Square Error (MSE) | 1.70 | 0.26 |
| Mean Absolute Percentage Error (MAPE) | 19.26 | 11.20 |
| Mean Error (ME) | 0.54 | 0.00 |
| Statistics for Seats Occupied ± One Seat | | |
| Mean Square Error (MSE) | 1.50 | 0.00 |
| Mean Absolute Percentage Error (MAPE) | 12.31 | 0.02 |
| Mean Error (ME) | 0.42 | 0.00 |

*Note*. The second set of statistics, Statistics for Seats Occupied ± One seat, refer to errors within one seat of the actual number of seats occupied. For example, if the actual number of seats occupied was 5 and the predicted number of seats occupied was 4 or 6 then the error would be 0, not 1, for this set of statistics.

### Data Mining

The data mining approach first forecasts if an hour at a game table will be empty or occupied. This is done using a gradient boosting method. Then, for hours marked as occupied, the number of seats occupied is forecast using random forests. We will discuss these two elements separately.

**Classification of empty or occupied seats.**

A target factor variable for the training data was set as Empty or Occupied depending on whether the actual Seats Occupied variable was zero or greater than zero. The set of predictor variables for this binary target variable was, in alphabetical order:

- Friday
- Game Code, MIN-BAC
- Hour of Day (12:00am through 11:00am)
- Hour of Day, lagged 1 hour
- Saturday
- Seats Occupied, lagged 1 hour
- Seats Occupied, lagged 1 week (Same Day of Week and Hour of Day lagged one week)
- Seats Occupied, lagged 2 hours
- Sunday

A gradient boosting method was employed to predict whether an hour at a game table was empty or occupied. The misclassification rate was 0.11 as reported in the Value for Data Mining Approach column of Table 5. The relative influence of predictors is shown in Figure 4. Seats Occupied, lagged 1 week clearly is the most influential predictor variable.

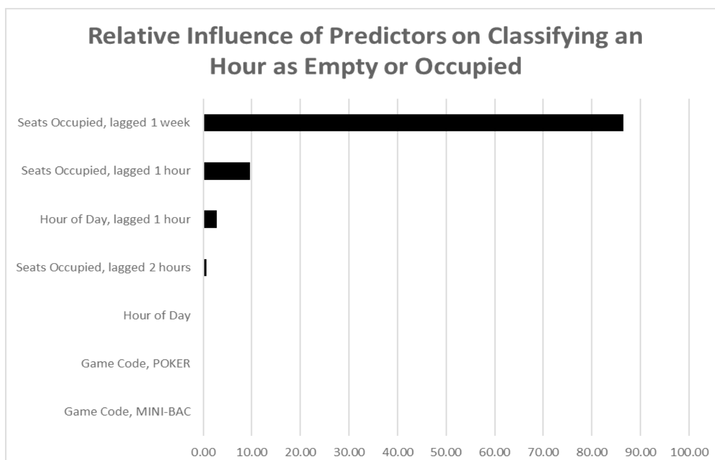


*Figure 4*. Relative influence of predictors on classifying an hour at a game table as empty or occupied.

**Forecasting the number of occupied seats for non-empty hours.**

The training data were filtered to remove those observations with zero occupied seats. This resulted in 1000 seven-day data sets with observations having non-zero occupied seats. The predictor variables used in the forecasting model were, in alphabetical order:

- Day of Week (Sunday through Saturday)
- Game Code (MIN-BAC or POKER)
- Hour of Day (12:00am through 11:00pm)
- Hour of Day, lagged 1 hour
- Seats Occupied, lagged 1 hour
- Seats Occupied, lagged 1 week (Same Day of Week and Hour of Day lagged one week)
- Seats Occupied, lagged 2 hours

A random forest model was applied to the data. The relative influence of the predictor variables to predicting seats occupied can be seen in Figure 5.

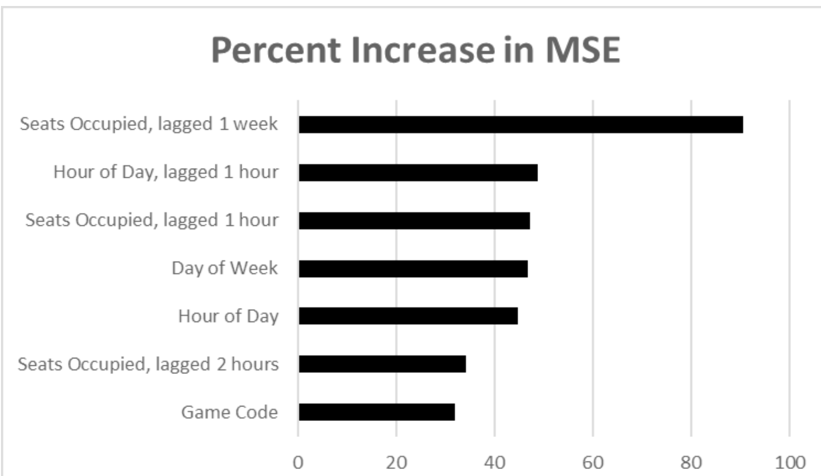


*Figure 5*. Percent increase of mean squared error for each predictor variable used in the random forest model. Percent increase in MSE is the most robust and informative measure of importance. It is the increase in MSE of predictions as a result of variable j being permuted (values randomly shuffled). Larger values indicate more importance to the forecast (Welling, 2015)

Seats Occupied, lagged 1 week again was the most influential predictor variable. Other reported variables show decreased influence on the prediction. The last variable, Game Code, is used to distinguish mini-baccarat from poker.

**Comparing Croston's Results to Data Mining Results**

Examining the rightmost two columns of Table 5, it can be seen that the data mining approach yields better results than does the Croston approach for all statistics. In particular, the Statistics for Seats Occupied ± One Seat area reports an MSE (mean squared error) of 1.50 for the Croston approach and an MSE of virtually zero for the data mining approach. Typically, MSE is used to compare the effectiveness of two or more techniques. A MAPE (mean absolute percentage error) of 12.31 is reported for the Croston approach and a MAPE of 0.02 is reported for the data mining approach. This is a considerable difference. Although MSE is the preferred statistic when comparing two approaches, MAPE is more readily understood by non-statisticians. Lastly, the Croston approach exhibits an ME (mean error) of 0.42 while the data mining approach exhibits ME of 0.00.

**Conclusions**

Although this research used simulated demand data, it shows that a two-stage data mining approach to forecasting demand may work better than Croston's method at forecasting game table demand. This observation is useful to casino management seeking a way to forecast game table traffic in order to plan staffing levels and to plan opening or closing game tables at different hours throughout the day. The data mining forecast can be used in place of Croston's forecast for the mixed integer linear programming formulation for assigning tables to betting limits reported in Chen, Tsai, and McCain (2012).

This study was performed using the open source R statistical computing language thereby avoiding the financial cost of using commercial software. This makes the use of the methods described in this paper more accessible to casino management.

## References


Breiman, L. & Cutler, A. (n.d.). Retrieved from www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.

Callegaro, A. (2010). Forecasting Methods for Spare Parts Demand.

Chen, M. Tsai, H., & McCain, S. (2012). A revenue management model for casino game tables. *Cornell Hospitality Quarterly*, 53(2), 144-153.

Croston, J. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23(3), 289-303.

Decision tree advantages (n.d.). Retrieved from https://en.wikipedia.org/wiki/Decision_tree_learning#Decision_tree_advantages.

Kandananoud, K. (2012). A Comparison of Various Forecasting Methods for Autocorrelated Time Series. *International Journal of Engineering Business Management*, 4.

Liaw, A. & Wiener, M., (2002). Classification and Regression by randomForest. *R News*, 2/3, December 2002.

McCabe, B. P. M., & Martin, G. M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting*, 21(2), 315-330.

Rao, A. V. (1973). A comment on: forecasting and stock control for intermittent demands. *Journal of the Operational Research Society*, 24(4), 639-640.

Scikit-learn (n.d.). Retrieved from http://scikit-learn.org/stable/modules/ensemble.html.

Snyder, R. D., Ord, J. K., & Beaumont, A. (2012). Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting*, 28(2), 485-496.

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, 21(2), 303-314.

Welling, S. H. (2015). Retrieved from http://stats.stackexchange.com/questions/162465/in-a-random-forest-is-larger-incmse-better-or-worse.

## Appendix A - Link to Code and Supporting Workbook for Demand Simulator

The following is a link to a zip file containing R code and a supporting Excel workbook used to generate training and testing data sets similar to those used in this study.

https://drive.google.com/file/d/0B3p-Oi0bCEoYZEJvMXhrWUxlU3M/view?usp=sharing

## Appendix B - Willemain and Desaultels Method to Construct Autocorrelated Uniform Random Numbers

Let $V_n$ be an independent and identically distributed uniform random number over the interval (0,1), F(X) be the cumulative distribution of the sum of a U(0,1) number and a U(0,c) number, and c be any positive number.

$X_1 = V_0 + cV_1$

$X_n = F(X_{n-1}) + cV_n$ , n>1 (for positive correlation).

The sequence of values $U_n = F(X_n)$ is autocorrelated U(0,1).

$F(X) = 1 - .5(1 + c - X)^2 / c$ X in the interval [1,1 + c] and $c \geq 1$.

Willemain and Desautels (1993) report a different form of F(X) when c is less than unity.