Report on BCLC retailer wins, by Prof. H. Joe, June 8, 2009

Summary

In this report, the analysis refers to win data for prizes exceeding $1000 in the period August 2007 to May 2009. 'Retailer win rate' means 'win rate by BCLC retailers' or the percentage of prizes in a category won by BCLC retailers.

- For non-Keno and non-Scratch & Win games, the retailer win rate could be explained by statistical variations and a higher spending rate among BCLC retailers. Sports Action wins were not examined separately because they do not depend on chance alone. Scratch & Win consists of over 30 different games with different wager amounts and different odds of winning prizes over $1000, so that analysis based on aggregate wins in Scratch & Win is not meaningful.
- For Keno, an analysis is more complicated because of the variety of possible combinations by which one can achieve a large prize. The seven win categories of matching 10 in pick 10, matching 9 in pick 10, matching 9 in pick 9, matching 8 in pick 9, matching 8 in pick 8, matching 7 in pick 7, and matching 6 in pick 6 all have a prize exceeding $1000 for any wager amount with or without the bonus option. The Ipsos Reid surveys still do not measure the frequency at which BCLC retailers and the general public play pick6, pick7, pick8, pick9, pick10 in Keno. Based on some rough adjustments from the Ipsos Reid surveys for Keno to estimate the frequency of playing in different Keno pick categories, the retailer win rates are within statistical variations given the higher spending rate.
- The win count totals are not high and it would take more (years of) data to have more power to detect if any observed retailer win rate is too high.

Recommendations for future

- In order to have a better analysis for Keno wins as the number of wins above $1000 gets higher in the next few years, it is crucial that the Ipsos-Reid survey have some questions on the frequency of play for each of pick6, pick7, pick8, pick9, pick10. The additional questions should be: "In the past year, how many times have you played Keno with pick i?" for $i = 6, 7, 8, 9, 10$. The responses to these 5 questions should be summarized into averages (one average for retailers, and one average for the general public; the averages should include the responses that are zero).
- The tracking of wins above $1000 does not really have much ability to detect fraudulent claims if there are any. Since you have names of each person with a prize above $1000, the additional analysis that should be done is to *flag* each person with more than one prize above $1000, and do additional checks if any person has won two or more prizes that are much larger than $1000.

Documents / information received

Count of number of BCLC retail employees : Additional_Data.doc
Reports of Ipsos-Reid polls : BCLC_Retailer_and_Player_Tracking_Wave_3_Final.pdf,
BCLC_Retailer_and_Player_Tracking_Wave_4_Final.pdf, BCLC_Retailer_and_Player_Tracking_Wave_5_Final.pdf
major prizes by lottery game : Major_Claims_Stats.xls
Keno win data: summary_keno_apr2009.xls

Computations

Let R_p = ratio of (participation rate for BCLC retail employee) to (participation rate for other adults). Let R_s = ratio of (average spent per game per participating BCLC retail employee) to (average spent per game for other participating adults). Let P = (size of BCLC retail employee subpopulation as a percentage of the BC adult population).

A key quantity in the analysis of whether retailers win "too much" is the product $w = P \times R_p \times R_s$. R_p, R_s are estimated by $\widehat{R}_p, \widehat{R}_s$ based on the Ipsos-Reid surveys.

For a win category in Keno such as matching 6 in pick 6 (Keno6/6), the bonus and wager amount are not relevant, and R_s should be (number of times the pick category is played in past year per BCLC retail employee) to (number of times the pick category is played in past year per adult member of general public). This definition of R_s is the same for other lottery games which do not have a variable wager amount.

Based on chance, the expected win rate for the retailer subpopulation for any prize category in a lottery game is w.

Let n_t be the total number of wins in a category, and let n_r be the total number of these wins claimed by BCLC retail employees. The point estimate of the retailer win rate w is $\widehat{w} = n_r/n_t$. To match up with an estimate based on the population size for BCLC retail employees and the ratio of the amounts spent by them relative to the general population, a confidence interval for w can be computed based on n_t and n_r.

Consider n_r as a realization of a Binomial(n_t, w) random variable. A $100(1-\alpha)\%$ confidence interval for w is:

$$\left\{w : \sum_{i \geq n_w} \binom{n_t}{i} w^i (1-w)^{n_t - i} \geq \alpha/2 \text{ and } \sum_{i \leq n_w} \binom{n_t}{i} w^i (1-w)^{n_t - i} \geq \alpha/2\right\}.$$

The interval can be computed by iteratively solving two equations, using code written in a statistical software such as R (http://www.r-project.org). We will use this interval with $\alpha = 0.01$ to get 99% confidence intervals.

A summary table of 99% confidence intervals, based on the above method, is given below.

Summary table with 99% confidence intervals (w_{LB}, w_{UB}) for w= retailer win rate						
game	#win	#rwin	w_{LB}	w_{UB}	$\widehat{w}$	$P\widehat{R}_p\widehat{R}_s$‡ waves 5,4,3,2
Lotto649	2610	25	0.005	0.016	0.010	0.015, 0.013, 0.009, 0.014
BC49	179	2	0.001	0.51	0.011	0.016, 0.010, 0.008, 0.010
extra	4165	58	0.010	0.019	0.014	0.018, 0.011, 0.008, 0.017
super7	624	8	0.004	0.029	0.013	0.018, 0.015, 0.011, 0.020
payday	10	1	0.001	0.54	0.100	0.024
pacific holdem	250	7	0.008	0.067	0.028	0.069
super7extra	1886	32	0.010	0.026	0.017	0.018, 0.015, 0.011, 0.020
keno	4910	80	0.012	0.022	0.016	0.055, 0.045, 0.040, 0.045
S&W	648	18	0.013	0.047	0.027	0.012, 0.009, 0.011, 0.016 [ave.=0.012]
keno9/10	50	2	0.002	0.17	0.040	0.015
keno9/9	2	0	0.000	0.93	0.000	0.015
keno8/9	89	4	0.008	0.13	0.045	0.015
keno8/8	38	1	0.000	0.18	0.026	0.015
keno7/7	176	8	0.015	0.10	0.046	0.015
keno6/6	2383	43	0.012	0.026	0.018	0.015
keno(above 6 combined)	2738	58	0.015	0.029	0.021	0.015=$P \times 1.10 \times 1.63$

‡ $P\widehat{R}_p\widehat{R}_s$ values below w_{LB} indicate observed retail win rate is too high

‡ Values of $\widehat{R}_p, \widehat{R}_s$ based on Ipsos-Reid surveys

‡ $P = 0.86\%$ for BC is based the average of 30749/3465088 and 28952/3469018,
adult population size of 3,465,088 and retail staff size of 30749 in Feb 2008
adult population size of 3,469,018 and retail staff size of 28952 in May 2008

The following is an explanation for the last column of the above table For waves 5,4,3,2, the ratios R_p for Lotto6/49 based on the Ipsos Reid surveys are respectively $0.88/0.84 = 1.05$, $0.85/0.88 = 0.97$, $0.79/0.85 = 0.93$, $0.85/0.84 = 1.01$; and the ratios R_s are respectively $163/100 = 1.63$, $164/107 = 1.53$, $133/113 = 1.18$, $190/114 = 1.67$ (see pp. 20–21 of the wave 5 report). Then $P \times R_p \times R_s = 0.0086 \times 1.05 \times 1.63 = 0.015$ for wave 5, etc.

In the above table, except for in the rows of Keno9/10 and below, the entries for $\widehat{R}_p$ and $\widehat{R}_s$ were determined from pages 20 and 21 of BCLC_Retailer_and_Player_Tracking_Wave_5_Final.pdf. The values for Super7Extra were taken to be the same as Super7. The values for Pacific Holdem and Payday were mostly missing, and the last column for Payday is taken from the Wave 2 report.

For Keno, the chance of winning $1000 or more increases with the wager amount and with the bonus option. In the above table, for the line with Keno and #win=4910, all categories of pick1,...,pick10, bonus or not, and all wager amounts were combined, but the assumptions behind the use of the formula for the confidence intervals are not valid. The row is there for comparison only.

For Scratch & Win, there are other 30 different games at www.bclc.com, with wager amount that could be $1, $2, $3, $5, $10 and odds of winning prizes over $1000 that vary a lot. The assumptions behind the use of the formula for the confidence intervals are not valid and this is an explanation of why Table 1 seems to suggest that the retail win rate is too high for Scratch & Win. An explanation could be that the retail employees more frequently play the Scratch & Win with higher wager amounts and higher chance of a prize over $1000.

For the calculations in the rows of Keno9/10 to the end of the table, the values for $\widehat{R}_p$ and $\widehat{R}_s$ are not as reliable, because Ipsos Reid does not ask the appropriate questions for frequency of play for Keno pick 6 to pick 10. For participation rate for pick categories of 5 or more, I am using the averages from figures in the third and fourth bullet of page 46 of the wave 5 report to get $R_p = 1.10$; the average spend ratio is 1.85 based on pages 20 and 21, and this is downweighted to 1.63 based on the wager amounts on pages 52 and 53 (retailers are wagering more on average).