A Short Study of Interclub Race Results and Optimum Q value.

Steve Corona
March 14, 2018

Q Primer

As I was beginning to communicate this information I was reminded that there might be individuals relatively new to PHRF-LO who might need a little background information. In addition, any document should be able to stand on its own without requiring a lot of background reading so I will add some information that may have been previously included in similar studies. Those already onboard may skip ahead.

The "Q" factor is a variable in the PHRF-LO time on distance (TOD) to time on time (TOT) conversion formula that can be used to adjust the relative handicap between faster and slower rated boats. The "Q" factor does not change relative positions of individual boat handicaps but rather adjusts the global handicap differences.

The following three graphs show a single race result corrected with several different Q values.

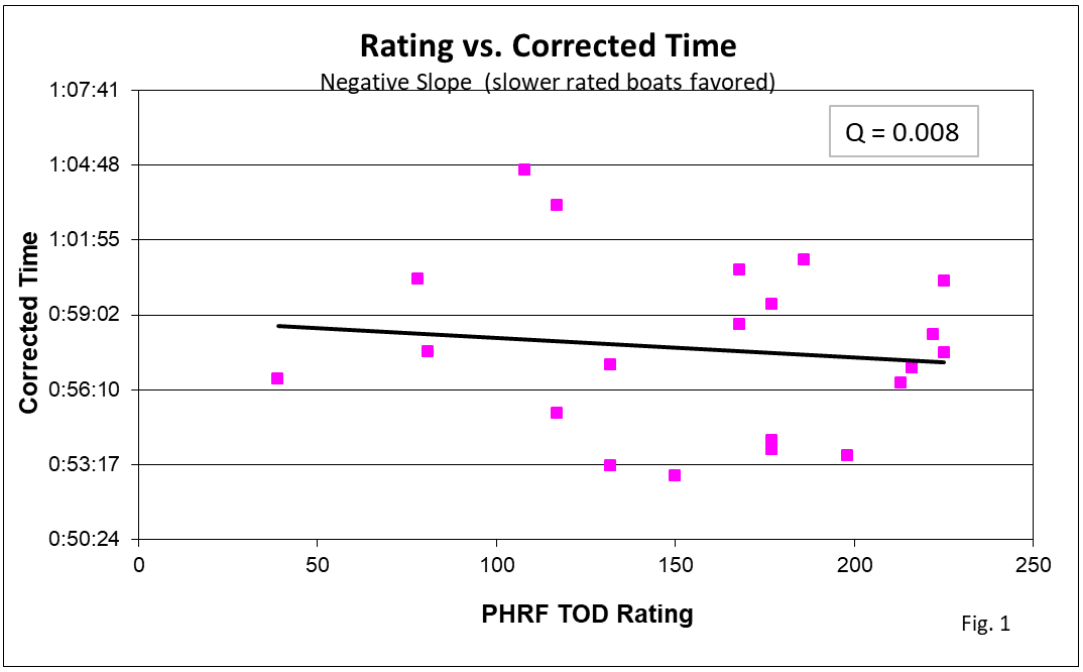


Figure #1 shows the corrected time of the individual boats plotted against their PHRF TOD rating. The race was corrected using Time on Time (TOT) with the PHRF-LO Q value of 0.008. At this Q value the linear regression of the data (dark straight line) has a negative slope indicating that the slower rated boats (higher PHRF TOD numbers) were slightly advantaged for this race. This is not unexpected because all races show a handicap bias in one direction or another. The goal is to average out the bias over time.

Figure #2 is the same race corrected with a Q value of 0.045. Note that the linear regression line has a positive slope indicating that the faster rated boats were slightly favored at a Q value of 0.045.

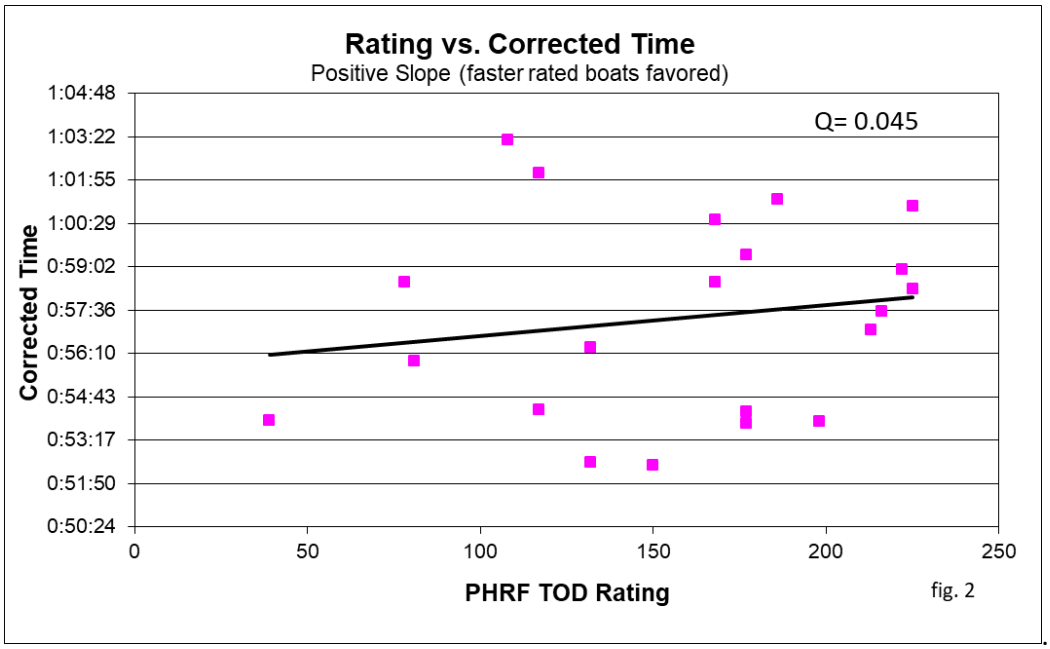
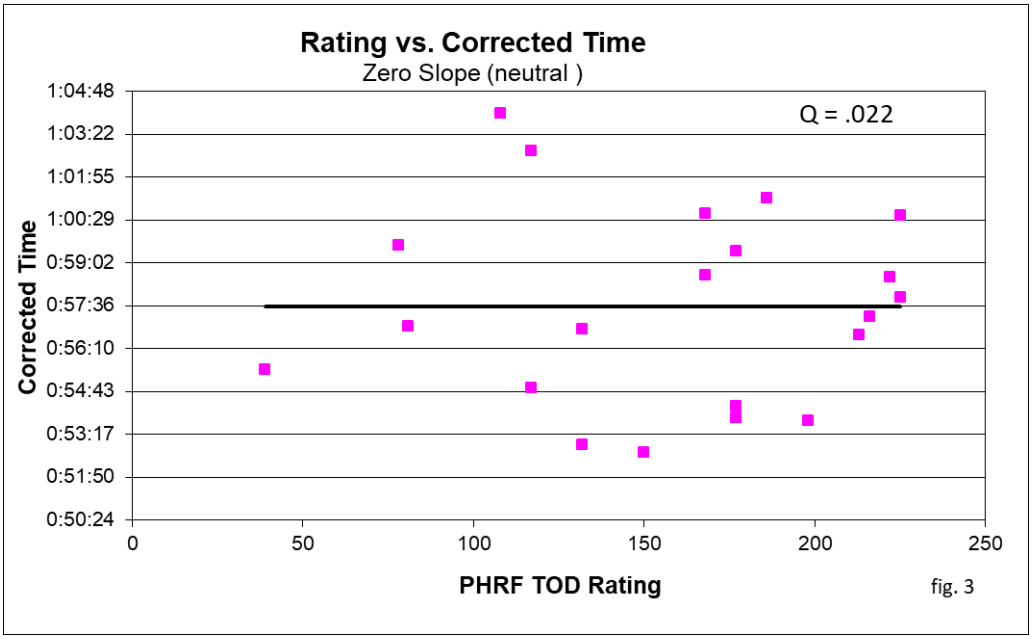


Figure #3 is again the same race except the Q value has been adjusted (Q = 0.022) to provide a zero slope of the regression line (no handicap bias).



A zero regression slope almost never occurs naturally. Either the slower or faster rated boats have an advantage in any race. What is important to remember in these examples is that a regression slope can be calculated and identified for every race but not every race will produce an optimum Q value.

Two analysis methods have been developed that can be used to calculate the optimum average Q value for a set of races: The first method (original) iteratively adjusts the Q value while correcting a race multiple times until a near zero regression slope is found as shown on figure #3. The individual optimum Q values from the set of races can then be statistically analyzed. Approximately one quarter to one third of races cannot be adjusted to produce a zero slope and are discarded for analysis purposes.

The second (newer) analysis method capitalizes on the fact that all races have a bias one way or the other (either positive or negative slope). In any data set if we can find a Q value that produces an equal number of races that are of positive and negative slopes, that Q value might also be called optimum.

In summary the first analysis method calculates the optimum Q individually race by race and the second method finds the Q value that appears to produce an equal number of positive and negative regression slopes in the whole data set.

RESULTS

Flying Sail Data 2013-2017 - Analysis Method #2

There are 593 races in the flying sail (FS) data set that are identified as "Interclub" and 4137 races that are not identified for a total of 4730 races. The races were corrected using a Q value of 0.008 (the present value used on Lake Ontario). The following chart (figure #4) shows the result using analysis method #2.

Interclub			Non-Interclub		
	Value	Percent		Value	Percent
Q =	0.008		Q =	0.008	
slope < 0	256	43%	slope < 0 =	2167	52%
slope = 0	0	0%	slope = 0	0	0%
slope > 0 =	337	57%	slope > 0 =	1970	48%
Mean =	0.001154		Mean =	0.000954	
Median =	0.000337		Median =	-0.000116	
# Races	593		# Races	4137	

fig.#4

At a Q value of 0.008 the analysis shows that, for FS interclub races, faster rated boats are advantaged 57% of the time (slope > 0) and for non-interclub races the slower rated boats are advantaged 52% of the time (slope < 0).

Flying Sail Data 2013-2017 - Analysis Method #1

For this analysis method, optimum Q (zero slope) was iteratively calculated for each race. For the FS data set, 458 interclub races were usable and 2698 non-interclub races. For the record, optimum Q could not be found for 23% of the interclub races and 35% of the non-interclub races.

Flying Sail Data 2013-2017			
Interclub Only		*Non-Interclub Only*	
Mean	0.013813	Mean	0.04096
Median	-0.03385	Median	-0.01685
Mode	-0.0991	Mode	-0.025
Standard Deviation	0.160394	Standard Deviation	0.193371
Range	1.1151	Range	1.1373
Minimum	-0.1415	Minimum	-0.1418
Maximum	0.9736	Maximum	0.9955
Sum	6.3264	Sum	110.5097
Count	458	Count	2698
Largest(1)	0.9736	Largest(1)	0.9955
Smallest(1)	-0.1415	Smallest(1)	-0.1418

fig.#5

Because of the non-linearity of Q, and the resulting skewed and truncated distribution produced, the statistical median is probably a more valid indicator of the optimum Q value than the mean. Both interclub and non-interclub data indicated a lower optimum Q value than the current Q value of 0.008. The interclub data shows a greater change than the non-interclub data. This is consistent with the results from analysis method #2.

Non-Flying Sail Data 2013-2017 - Analysis Method #2

In the non-flying sail (NFS) data set 255 races were identified as interclub races and 4307 races as non-interclub races. All races were corrected using a Q value of 0.008. The following chart (fig.#6) shows the results of the regression slope analysis using analysis method #2.

Interclub			**Non-Interclub**		
	Value	**Percent**		**Value**	**Percent**
Q =	0.008		Q =	0.008	
slope < 0	89	34.9%	slope < 0	2067	48.0%
slope = 0	0	0.0%	slope = 0	1	0.0%
slope > 0	166	65.1%	slope > 0	2239	52.0%
Mean =	0.000399		Mean =	0.000007	
Median =	0.000451		Median =	0.000104	
# Races =	255		# Races =	4307	

fig.#6

The analysis shows that in NFS interclub races, at a Q value of 0.008, the faster rated boats are favored 65% of the time. For non-interclub races the faster rated boats are favored 52% of the time.

Non-Flying Sail Data 2013-2017 - Analysis Method #1

Analysis method #1 calculates (where possible) the optimum Q (zero slope value) for each race. For the interclub NFS data 209 races were usable and for the non-interclub data 3244 races. Optimum Q could not be found for 18% of the interclub and 25% of the non-interclub races. The following chart (fig.#7) displays the results.

Non-Flying Sail Data 2013-2017			
NFS Interclub Data		*NFS Non-Interclub Data*	
Mean	0.0013	Mean	0.0266
Median	-0.0486	Median	-0.0297
Mode	-0.0665	Mode	0.0080
Standard Deviation	0.174355	Standard Deviation	0.182230
Range	1.1239	Range	1.1375
Minimum	-0.1409	Minimum	-0.142
Maximum	0.983	Maximum	0.9955
Sum	0.2779	Sum	86.2656
Count	209	Count	3244
Largest(1)	0.983	Largest(1)	0.9955
Smallest(1)	-0.1409	Smallest(1)	-0.142

fig.#7

The predicted median optimum Q values of both the interclub and non-interclub NFS data are lower than the present Q value of 0.008. The interclub data clearly predicts a lower optimum Q value than the non-interclub data (faster rated boats advantaged). The results of analysis method #1 is consistent with the results obtained with analysis method #2.

SUMMARY AND CONCLUSIONS

1. Both the flying sail (FS) and non-flying sail (NFS) interclub data appears to show that the faster boats in any division are advantaged at a Q value of 0.008 .

2. The FS and NFS non-interclub data (the great majority of races) are relatively neutral (no handicap bias at Q = 0.008).

3. The simplicity of analysis method #2, and the fact that all races are considered, increases my confidence in this method. Analysis method #1 is a good backup and confirmation.