SAW 47 Working Paper 6 (TOR 2b) – Treatment of Zeros

October 30, 2006

The treatment of "zero" observations in the Summer Flounder ADAPT VPA calibration

by
Mark Terceiro

Introduction

The issue of how to treat "zero" observations in ADAPT VPA calibrations was addressed in a previous Southern Demersal Working Group (SDWG) working paper used in preparing the 2004 summer flounder assessment (SDWG 2003; beginning on page 8). That work responded to the 2002 SAW 35 (NEFSC 2002) summer flounder assessment Research Recommendation: *Explore the sensitivity of the VPA calibration to the addition of 1 and/or a small constant to values of survey series with "true zeros."* This recommendation stemmed from the nature of the ADAPT VPA calibration (tuning) algorithm, which includes natural logarithm (ln) transformation (i.e., assumption of a lognormal error distribution) of the input survey abundance indices prior to calibration. Some of the tuning series in the assessment include several "true zero" observations (as contrasted with years for which no sampling was performed) in their time series. Since "zeros" are treated as missing values in the ADAPT computations, a possible solution would be to add a constant to every value in these series to enable use of these "true zeros" as observations.

In the 2002 (NEFSC 2002) and 2003 (Terceiro 2003a) summer flounder assessments, the addition of the constant value of 1 was made for five age 0 recruitment indices: the MA DMF Seine, CT DEP fall trawl, RI DFW fall trawl, RI DFW monthly trawl, and DE DFW 16 foot bay trawl survey series (note that the latter series was not included in the final ADAPT VPA tuning configuration). No constant was added to survey series with "zero" observations for other age classes. The choice of the value of 1 as the additive constant was based on recommendations from statistical texts (e.g., Snedecor and Cochran 1967, Sokal and Rohlf 1981) for the ln-transformation of data.

Berry (1987) provides guidance on the objective selection of the appropriate value of the additive constant based on the statistical properties (skewness and kurtosis) of data series to be ln-transformed. Briefly, the method consists of 1) addition of a range of constants from very large (e.g., 100) to very small (e.g., 0.0001) to the original values in the series, 2) ln-transformation of the modified series, 3) calculation of the skewness and kurtosis of the modified series, and 4) summation of the absolute value of the skewness and kurtosis (providing the statistic *g*) of the modified series. The additive constant that minimizes *g* for a given series of data is the one that best minimizes the effect of outliers and normalizes residuals from the lognormal error distribution, hence best adhering to the assumption of the lognormal distribution. Work using the procedures suggested by Berry (1987) with recreational fishery catch rates as indices of abundance indicated that the additive constant of 1 was an appropriate value for those data, typically with values between zero and 50 (Terceiro 2003b).

The SDWG (2003) work applied the method suggested by Berry (1987) to summer flounder age 0 surveys with "zero" observations. Of the five age 0 series with "zero" observations, the MA DMF series varies between 0 and 70, while the other four series contained small values that varied between 0 and 1. The 2003 work (SDWG 2003) found that for the MA DMF series, the additive constant of 1 minimized the value of g. For the other four series, g was minimized by small values of the additive constant ranging from 0.001 to 0.1, with an "average" best additive constant of 0.1. The SDWG (2003) therefore recommended use of the revised, varying (1 or 0.1) additive constants in future assessments, and this revision was made in the 2004-2006 assessment, for age 0 survey series only. No constant was added for survey series of other age classes, pending further research.

Recently, the 2006 assessment of summer flounder (Terceiro 2006a) was subject to a NMFS Office of Science and Technology (S&T) Peer Review (Methot 2006). Among the recommendations made by the S&T Peer Review panel was the following:

The Panel finds that one immediate modification of the VPA is justifiable and reduces the retrospective pattern in stock size during 2003-2005. The VPA model currently treats survey observations of zero as missing values. An observation of zero for a particular age of fish in a particular survey year does not mean that there are no fish of that age in the stock, only that the number of survey samples was not sufficient to detect any fish of that age. This VPA model, as with most assessment models, tunes to the logarithm of the survey observations so cannot explicitly deal with observations of zero. However, treating these zeroes as missing values can result in a bias because time periods of low abundance are underrepresented in the data input to the assessment model. In the case of summer flounder, the result may be an underestimate of the degree to which the stock has rebuilt since the low levels that occurred around 1990. The committee did not discuss this issue during the Sept 14-15 meeting, so is not prepared to present a definitive solution. An interim approach would use a small value in place of the zeroes. A value equal to one sixth of the smallest observed positive value would be reasonable until a more complete statistical solution can be developed.

As a result, a revised 2006 ADAPT VPA for summer flounder was developed for which the previous treatment of "zero" observations for age 0 indices was retained (additive constant of 1 for MA DMF seine survey, 0.1 for the CT DEP fall trawl, RI DFW fall trawl, RI DFW monthly trawl, and DE DFW 16 foot bay trawl surveys). For ages 1-7+ survey series with "zero" observations, a value equal to one-sixth of the minimum value in each series was used in place of the "zero" observations. Typically, the minimum non-zero value in these series was 0.01, and so the additive constant was 0.001667 (Terceiro 2006b).

Summer flounder 2006 ADAPT VPA

In this work, the Berry (1987) approach is applied to the summer flounder survey series for all ages with observed "zeros" to determine the best additive constant to use to remove these "zero" observations from the ADAPT VPA calibration data. Table 1 summarizes the statistical properties of the 24 survey series that were examined. The distributions of the surveys are characterized by non-zero values between 0.001 and 70, CVs that generally exceed 100%, positive skewness (long right hand tail), and significant kurtosis (high degree of peak, or contagion, near the mean). The proportion of "zeros" in the time series ranged from 1 of 31 = 3% (NEFSC Spring Age 3 index) to 13 of 28 = 46% (MA Fall 4).

Table 2 summarizes the results of the exercise for each group of age-specific indices. Values of *g* were minimized for constants between 0.001 and 100 (minimum values in ***bold italics***), for the age 0, 1, 2, 3, 4, and 5-7+ (aggregate) survey indices (number per tow or haul). There is no statistically significant correlation (Table 3) between the value of the additive constant that minimizes g and the statistical parameters listed in Table 1.

Age 0 Indices

For the five age 0 series, the *g* statistic was minimized by values of the additive constant ranging from 0.001667 to 1. The constant equated to one-sixth of the minimum non-zero observed value for 2 of the 5 series . The relationships between the additive constants and calculated values of *g* for the age 0 indices are shown in Figure 1.

Age 1 Indices

For the three age 1 series, the additive constant of 0.01 minimized the absolute value of *g*. The relationships between the additive constants and calculated values of *g* for the age 1 indices are shown in Figure 2.

Age 2 Indices

For the single age 2 series, the additive constant of 0.1 minimized the absolute value of *g*. The relationships between the additive constants and calculated values of *g* for the age 2 indices are shown in Figure 3.

Age 3 Indices

For the six age 3 series, the absolute value of the *g* statistic was minimized by values of the additive constant ranging from 0.001 to 100. The constant equated to one-sixth of the minimum non-zero observed value for 1 of the 6 series . The relationships between the additive constants and calculated values of *g* for the age 3 indices are shown in Figure 4.

Age 4 Indices

For the six age 4 series, the absolute value of the *g* statistic was minimized by values of the additive constant ranging from 0.001 to 1. The constant equated to one-sixth of the minimum non-zero observed value for 1 of the 6 series . The relationships between the additive constants and calculated values of *g* for the age 4 indices are shown in Figure 5.

Age 4/5-7+ Indices

For the three age 4/5-7+ series, the absolute value of the *g* statistic was minimized by values of the additive constant ranging from 0.001667 to 100. The constant equated to one-sixth of the minimum non-zero observed value for 1 of the 3 series . The relationships between the additive constants and calculated values of *g* for the age 4/5-7+ indices are shown in Figure 6.

Conclusion

There is no consistent pattern in the identification of the additive constant that minimizes the absolute value of Berry's (1987) *g* statistic. There is no strong relationship between the absolute magnitude of the index values, the length of the time series, the number of zeros, the magnitude of the smallest observed value, or any of the usual statistical moments of the series (mean, maximum, non-zero minimum, CV, skewness, kurtosis), and the value of the additive constant that minimizes *g*. Further, while the “one-sixth” of the minimum observed value was identified as the “best” additive constant in 5 of the 24 (21%) cases examined, this level is not high enough to justify this approach as a reliable rule-of-thumb. In fact, the additive constant of 0.01 was identified as “best” for a higher percentage of series (6 of 24 = 25%). Given the inability to identify a constant that consistently minimizes *g*, the best rule is to maintain the current approach of making no adjustment and continue to treat “zero” observations as “missing.”

References Cited

Berry DA 1987. Logarithmic transformations in ANOVA. Biometrics 43:439-456.

Methot R. 2006. Review of the 2006 Summer Flounder Assessment Update. Chair’s Report. NMFS Office of Science and Technology. 6 p.

Northeast Fisheries Science Center (NEFSC) 2002. Report of the 35th Northeast Regional Stock Assessment Workshop (35th SAW): SARC Consensus Summary of Assessments. NEFSC Ref Doc. 02-14. 259 p.

Snedecor GW, Cochran WG. 1967. Statistical methods (6th ed). Iowa State University Press. Ames IA. 593 p.

Sokal RR, Rohlf FJ. 1981. Biometry (2nd ed). W.H. Freeman and Company. New York, NY. 859 p.

Southern Demersal Working Group (SDWG). 2003. SAW Southern Demersal Working Group (WG) Responses to 2002 SAW 35 and 2003 Summer Flounder Assessment Research Recommendations (numbered as in the 2003 assessment): December 22, 2003. 16 p.

Terceiro M. 2003a. Stock assessment of summer flounder for 2003. NEFSC Ref Doc. 03-09. 179 p.

Terceiro M. 2003b. The statistical properties of recreational catch rate data for some fish stocks off the northeast U.S. coast. Fish Bull. 101(3): 653-672.

Terceiro M. 2006a. Stock assessment of summer flounder for 2006. NEFSC Ref Doc. 06-17. 119 p.

Terceiro M. 2006b. Summer flounder assessment and biological reference point update for 2006. 64 p.

Table 1 . Statistical properties of summer flounder ADAPT VPA survey calibration series with "zero" observations.

Survey Name	N Obs	N "Zeros"	Mean	Max	Non-zero Min	CV (%)	Skew	Kurt
Age 0 Indices								
RI Fall 0	26	2	0.130	0.550	0.01	118	1.422	1.212
RI Monthly 0	16	3	0.037	0.110	0.01	95	1.044	0.152
CT Fall 0	22	4	0.085	0.442	0.013	128	2.078	4.733
MA Seine 0	24	3	8.292	70.000	1	170	3.904	17.141
DE 30 0	15	1	0.487	2.280	0.02	133	1.874	3.463
Age 1 Indices								
MA Spring 1	28	2	0.260	1.770	0.025	140	3.105	11.161
MA Fall 1	28	2	0.761	2.907	0.011	109	1.280	0.686
RI Fall 1	26	1	0.535	2.470	0.05	120	2.197	4.724
Age 2 Indices								
MA Fall 2	28	2	0.759	2.235	0.047	85	0.884	-0.224
Age 3 Indices								
NEC Spring 3	31	1	0.302	1.020	0.01	99	0.865	-0.536
NEC Fall 3	24	2	0.168	0.660	0.01	111	1.076	0.334
MA Fall 3	28	2	0.132	0.756	0.010	132	2.086	5.191
RI Monthly 3	16	2	0.199	0.530	0.01	95	0.786	-0.916
NJ Trawl 3	18	3	0.340	1.280	0.01	112	1.141	0.828
DE 30 3	15	2	0.155	0.470	0.01	105	0.991	-0.137
Age 4 Indices								
NEC Spring 4	31	5	0.092	0.310	0.01	111	0.985	-0.404
NEC Fall 4	24	8	0.043	0.190	0.01	144	1.444	0.762
MA Spring 4	28	5	0.086	0.317	0.010	116	1.187	0.242
MA Fall 4	28	13	0.019	0.186	0.01	196	3.484	14.026
RI Fall 4	26	5	0.035	0.280	0.01	179	2.961	9.516
RI Monthly 4	16	4	0.060	0.240	0.01	122	1.257	0.856
Age 5-7+ Indices								
NEFSC Spring 5-7+	31	10	0.060	0.210	0.01	121	0.892	-0.793
NEFSC Winter 5-7+	15	1	0.803	2.600	0.01	106	0.698	-0.636
NJ Trawl 4-7+	18	4	0.172	0.810	0.01	129	1.715	2.946

Table 2. Values of the additive constants that minimize the statistic *g*. Values that are one-sixth the minimum observed in the series are in bold.

Survey Name	**Constant**
Age 0 Indices	
RI Fall 0	**0.001667**
RI Monthly 0	0.01
CT Fall 0	0.01
MA Seine 0	1
DE 30 0	**0.003333**
Age 1 Indices	
MA Spring 1	0.01
MA Fall 1	0.01
RI Fall 1	0.01
Age 2 Indices	
MA Fall 2	0.1
Age 3 Indices	
NEC Spring 3	0.01
NEC Fall 3	**0.001667**
MA Fall 3	0.001
RI Monthly 3	0.01
NJ Trawl 3	2
DE 30 3	100
Age 4 Indices	
NEC Spring 4	0.001
NEC Fall 4	0.1
MA Spring 4	1
MA Fall 4	0.001
RI Fall 4	**0.001667**
RI Monthly 4	0.1
Age 5-7+ Indices	
NEFSC Spring 5-7+	100
NEFSC Winter 5-7+	**0.001667**
NJ Trawl 4-7+	0.1

Table 3. Correlation analysis (value of *r*) of various statistical properties of the 24 summer flounder index series with "zero" observations. For n = 24, the critical value of *r* at the 0.05 significance level is about 0.4.

	N	Nzero	Mean	Max	Min	CV	Skew	Kurt	*g*
N	1.00								
Nzero	0.32	1.00							
Mean	0.01	-0.12	1.00						
Max	0.02	-0.08	1.00	1.00					
Min	0.18	-0.28	-0.01	-0.05	1.00				
CV	0.20	0.57	0.31	0.35	-0.19	1.00			
Skew	0.21	0.26	0.50	0.53	0.08	0.88	1.00		
Kurt	0.20	0.26	0.57	0.60	0.05	0.86	0.99	1.00	
g	-0.02	0.25	-0.08	-0.07	-0.12	-0.13	-0.24	-0.22	1.00

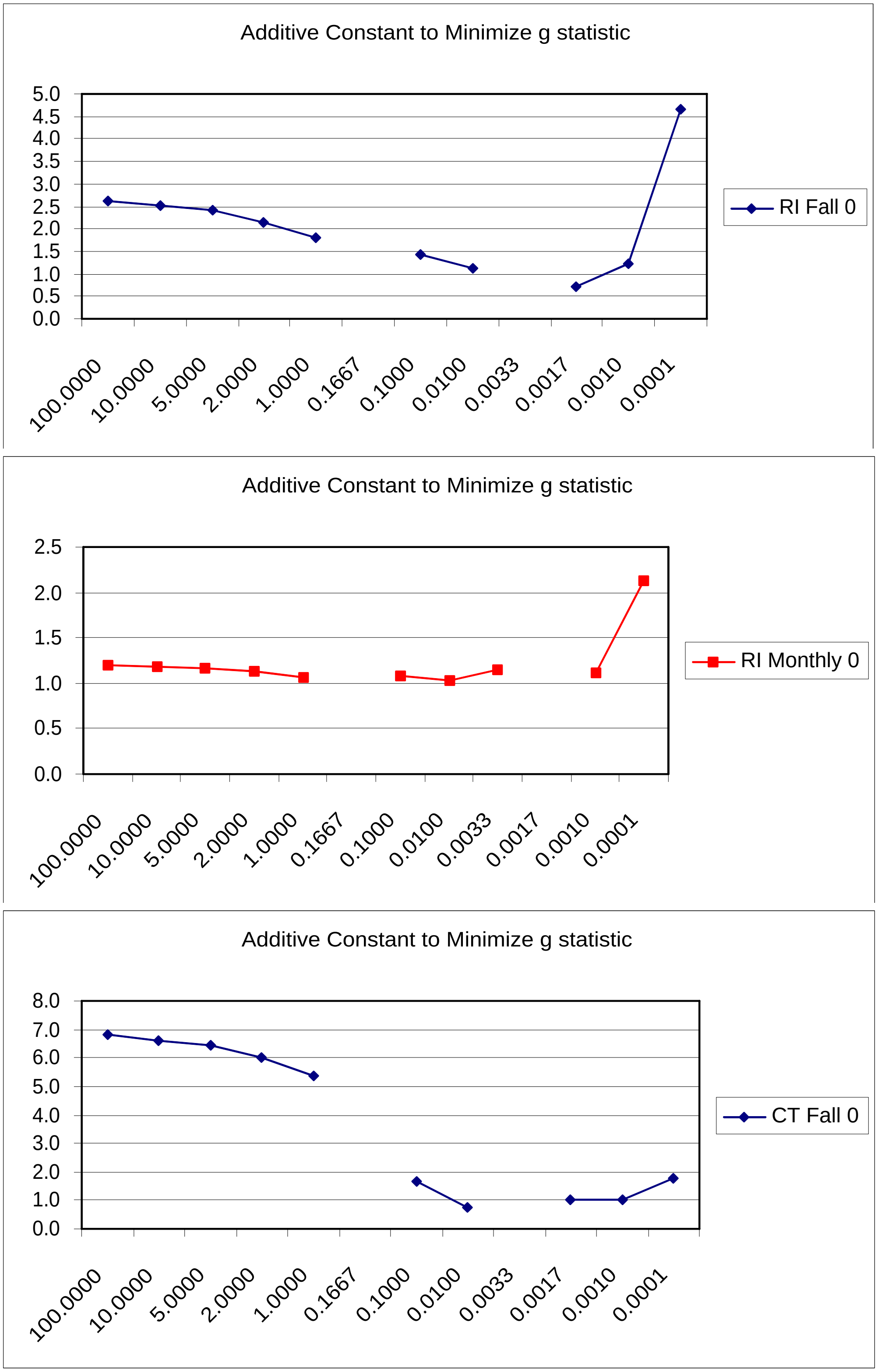

Additive Constant to Minimize g statistic
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1667
0.1000
0.0100
0.0033
0.0017
0.0010
0.0001
RI Fall 0
Additive Constant to Minimize g statistic
2.5
2.0
1.5
1.0
0.5
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1667
0.1000
0.0100
0.0033
0.0017
0.0010
0.0001
RI Monthly 0
Additive Constant to Minimize g statistic
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1667
0.1000
0.0100
0.0033
0.0017
0.0010
0.0001
CT Fall 0

Figure 1.

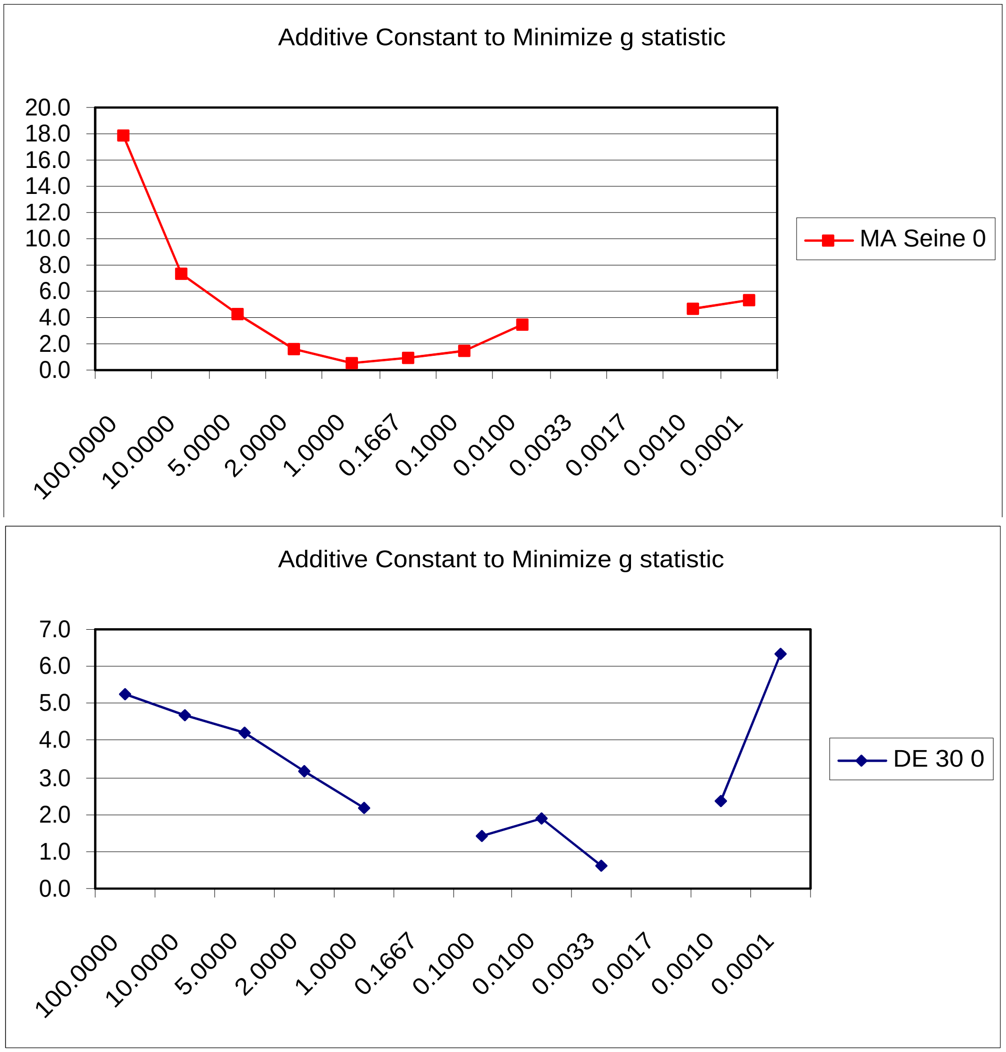


Figure 1 continued.

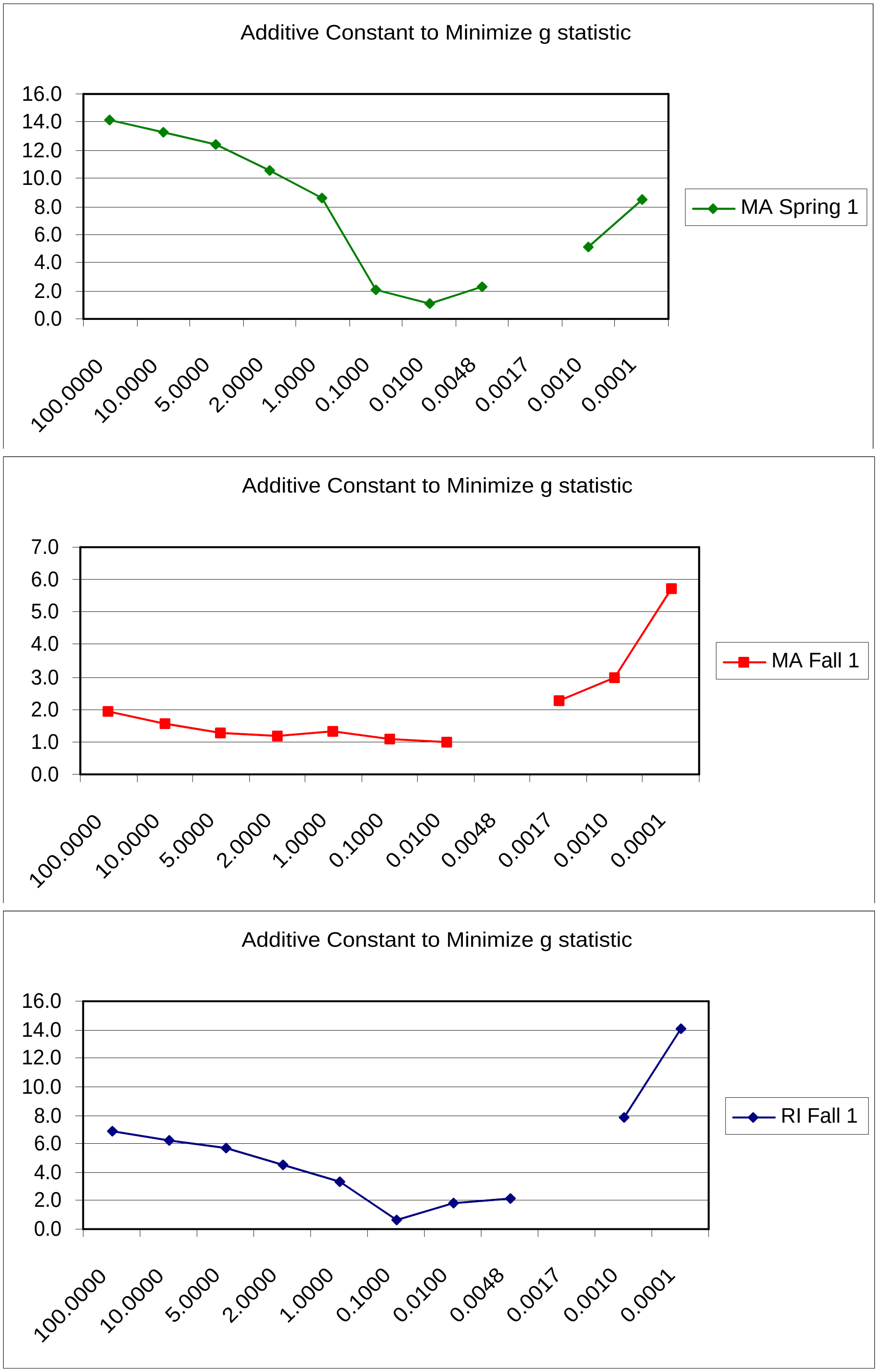

Additive Constant to Minimize g statistic
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0048
0.0017
0.0010
0.0001
MA Spring 1
Additive Constant to Minimize g statistic
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0048
0.0017
0.0010
0.0001
MA Fall 1
Additive Constant to Minimize g statistic
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0048
0.0017
0.0010
0.0001
RI Fall 1

Figure 2.

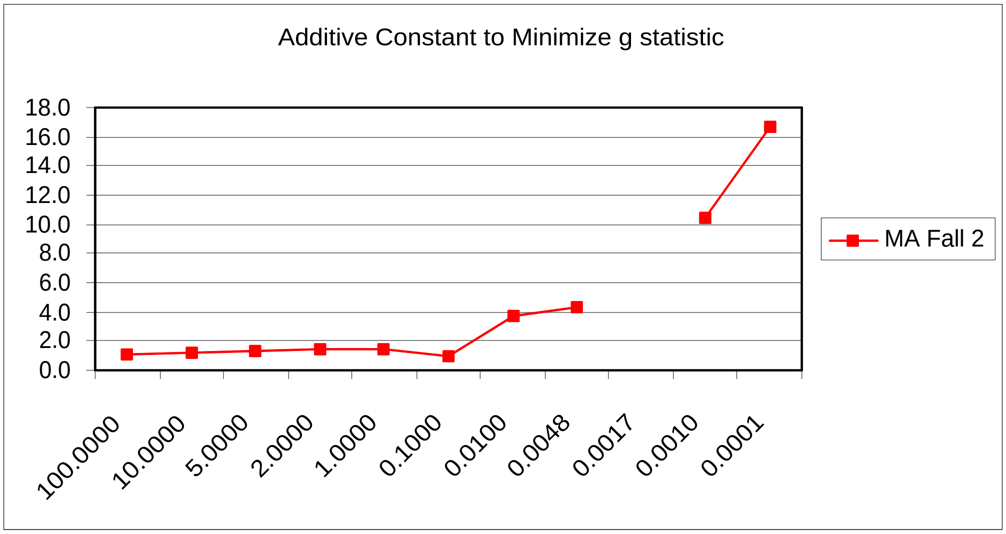

Additive Constant to Minimize g statistic
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0048
0.0017
0.0010
0.0001
MA Fall 2

Figure 3.

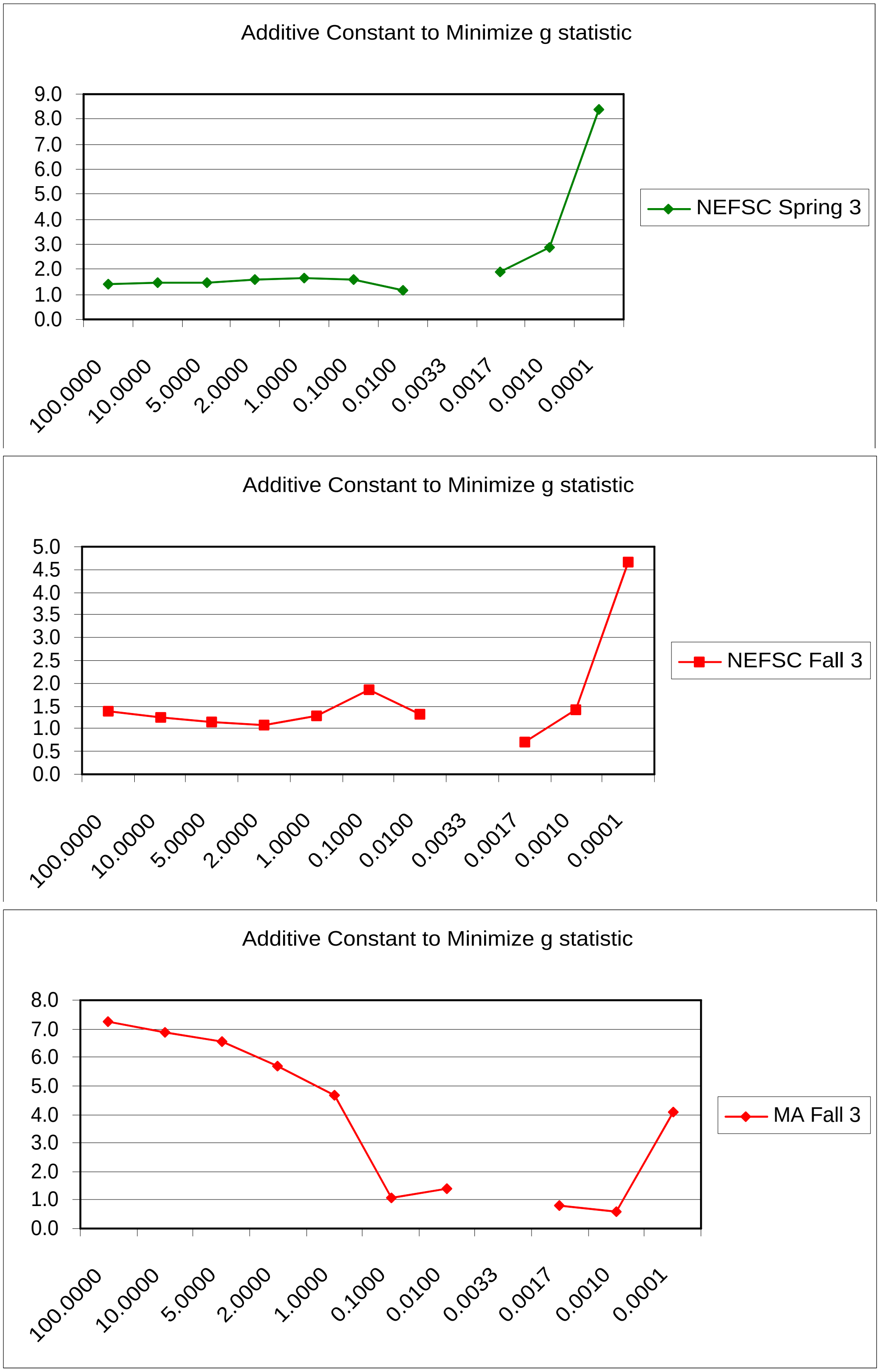

Additive Constant to Minimize g statistic
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0033
0.0017
0.0010
0.0001
NEFSC Spring 3
Additive Constant to Minimize g statistic
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0033
0.0017
0.0010
0.0001
NEFSC Fall 3
Additive Constant to Minimize g statistic
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0033
0.0017
0.0010
0.0001
MA Fall 3

Figure 4.

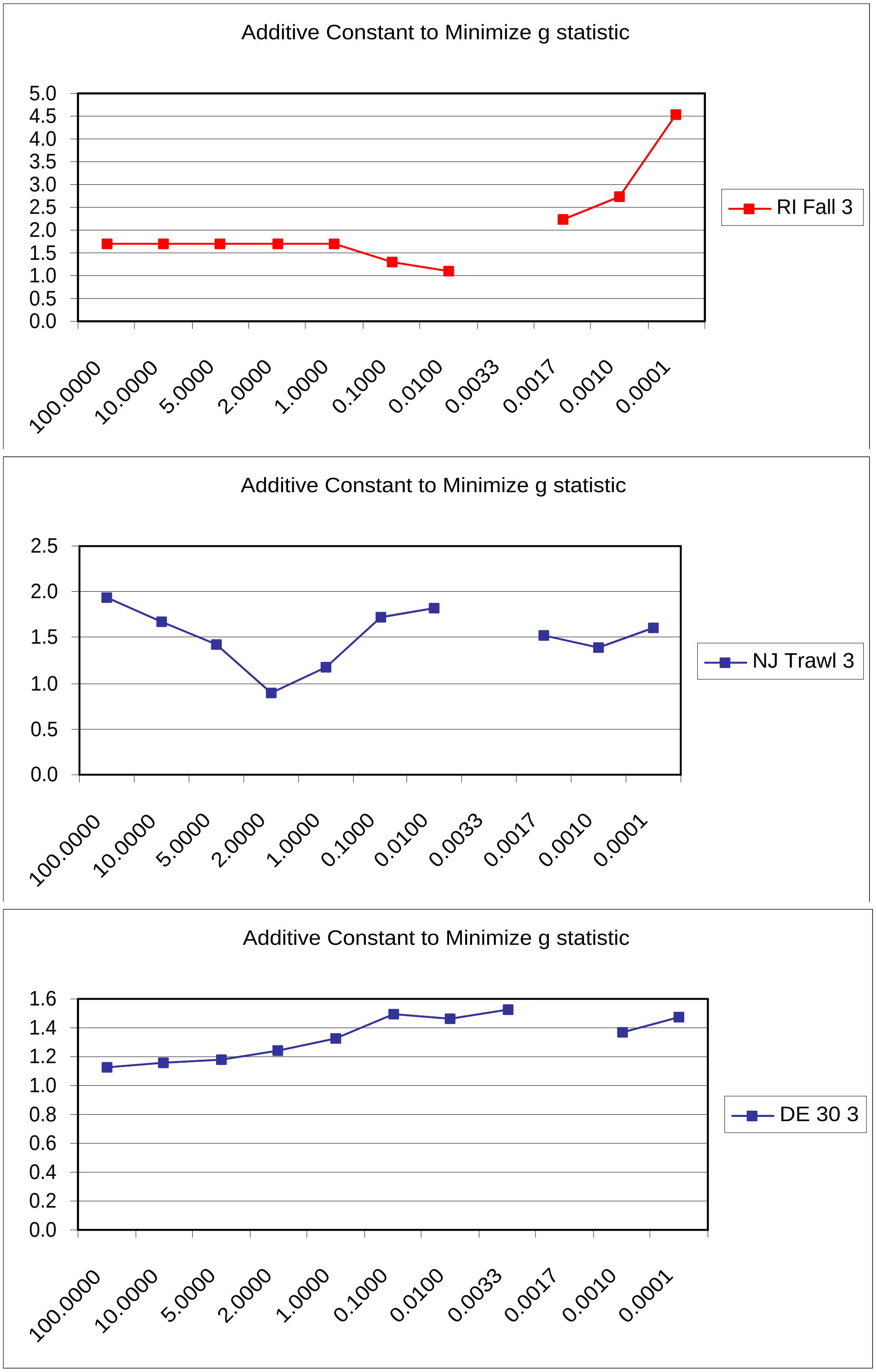
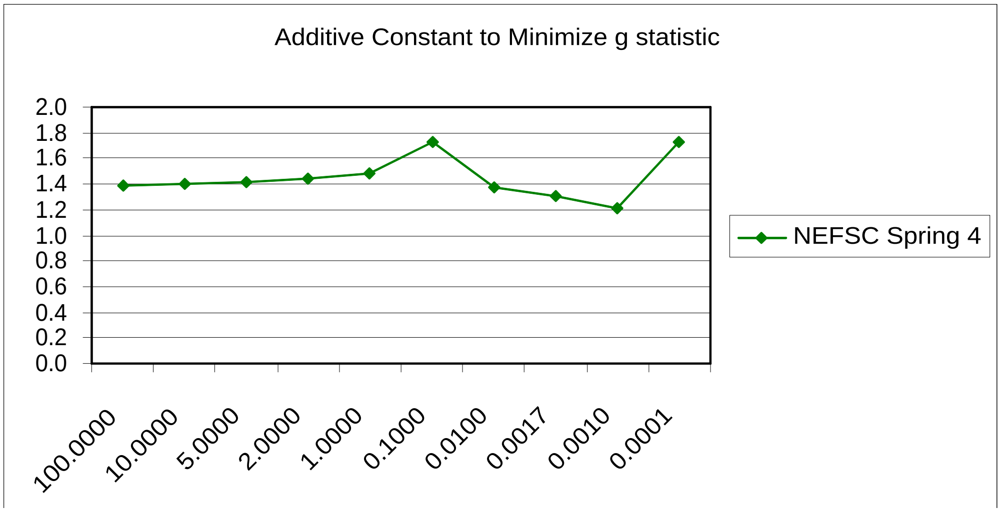
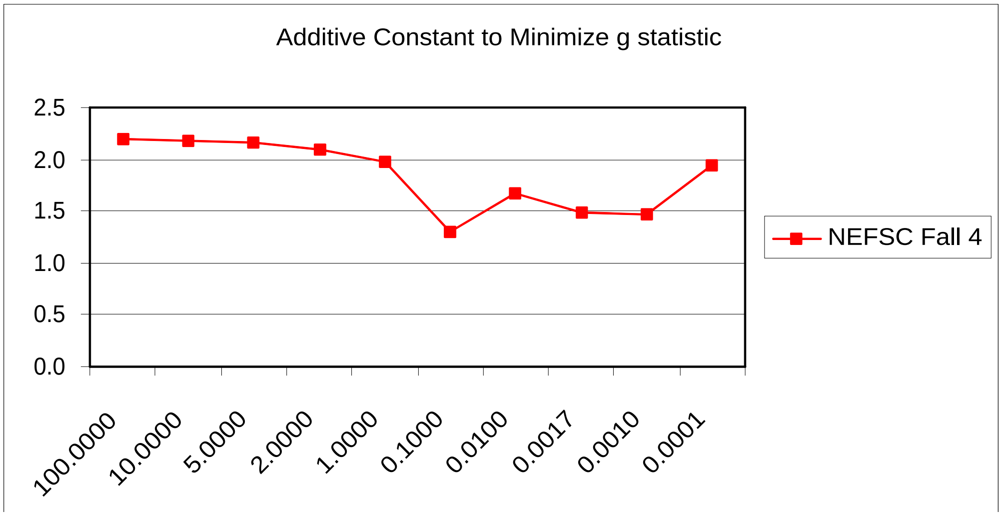


Figure 4 continued.


Additive Constant to Minimize g statistic
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0017
0.0010
0.0001
NEFSC Spring 4


Additive Constant to Minimize g statistic
2.5
2.0
1.5
1.0
0.5
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0017
0.0010
0.0001
NEFSC Fall 4

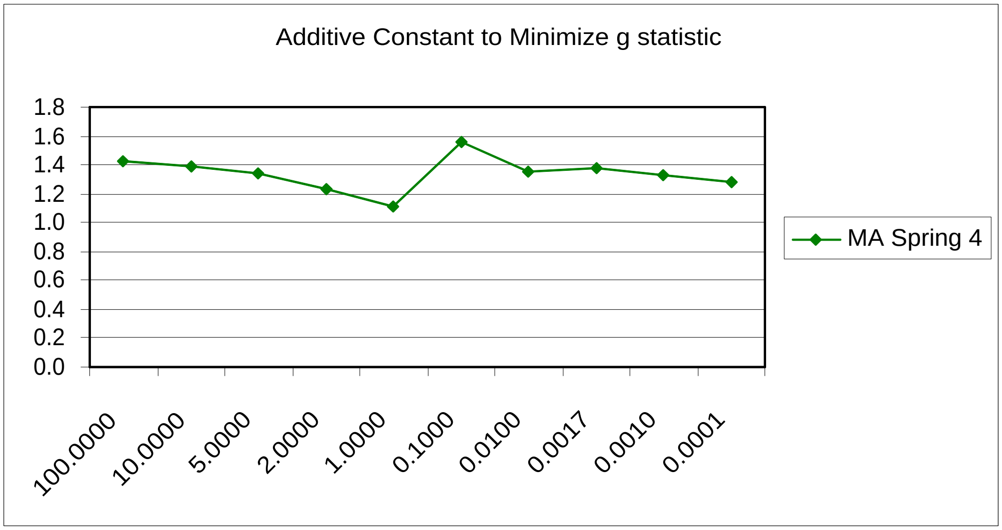
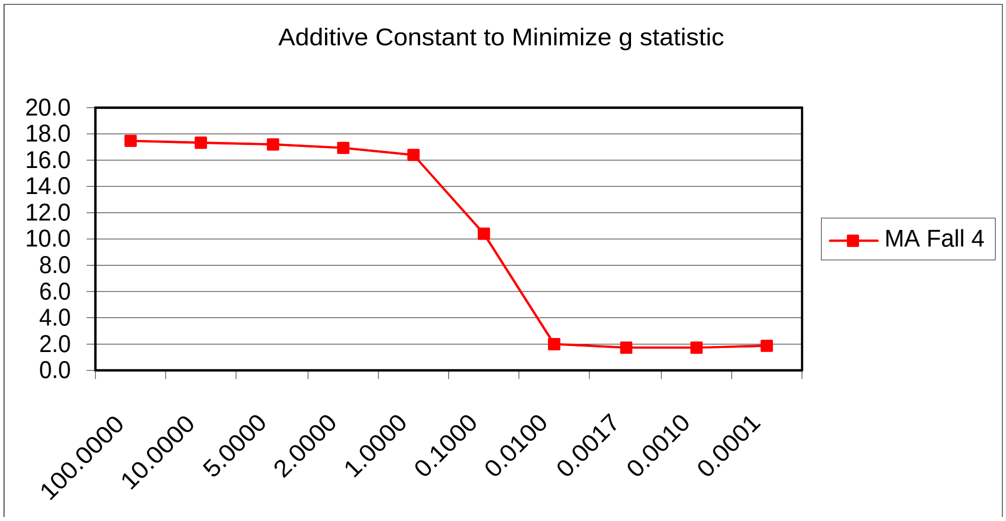
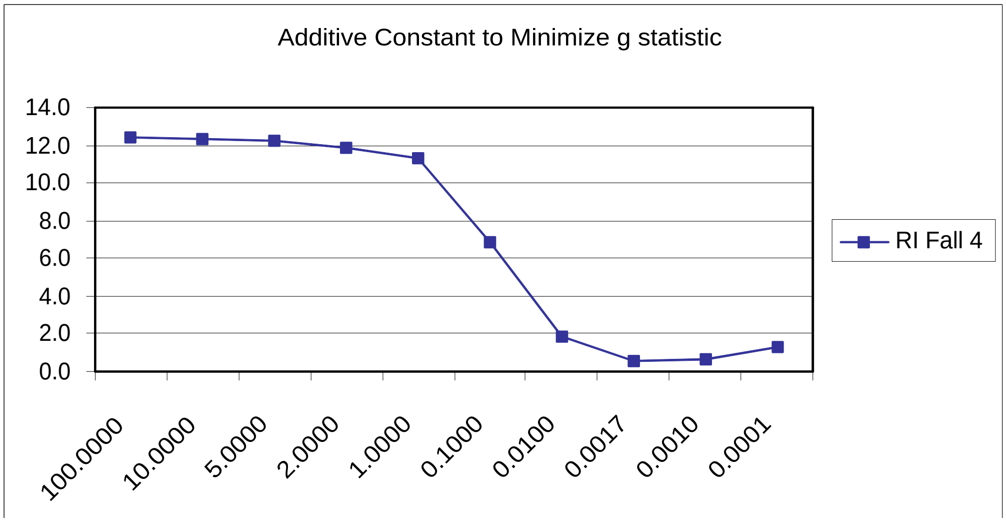

Additive Constant to Minimize g statistic
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0017
0.0010
0.0001
MA Spring 4

Figure 5.





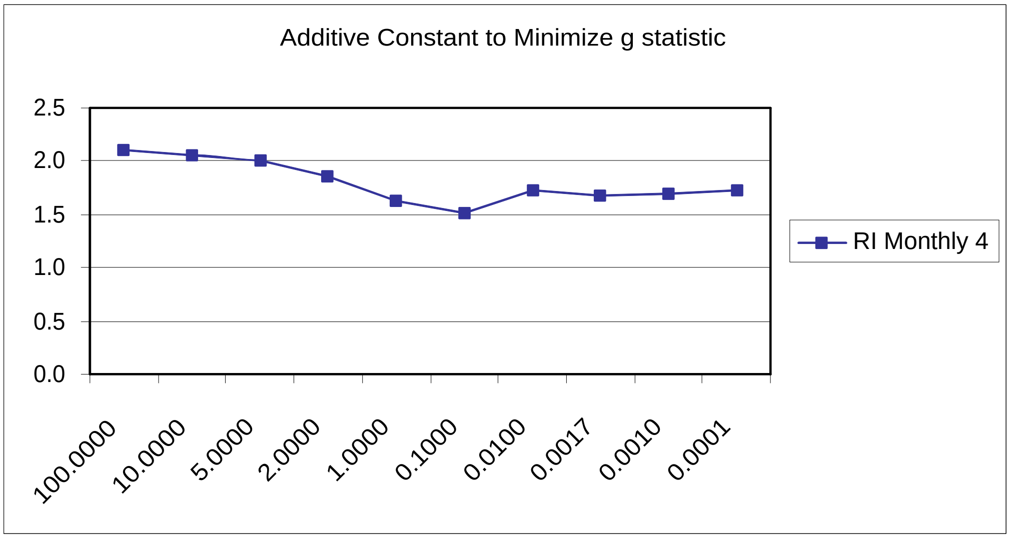


Figure 5 continued.

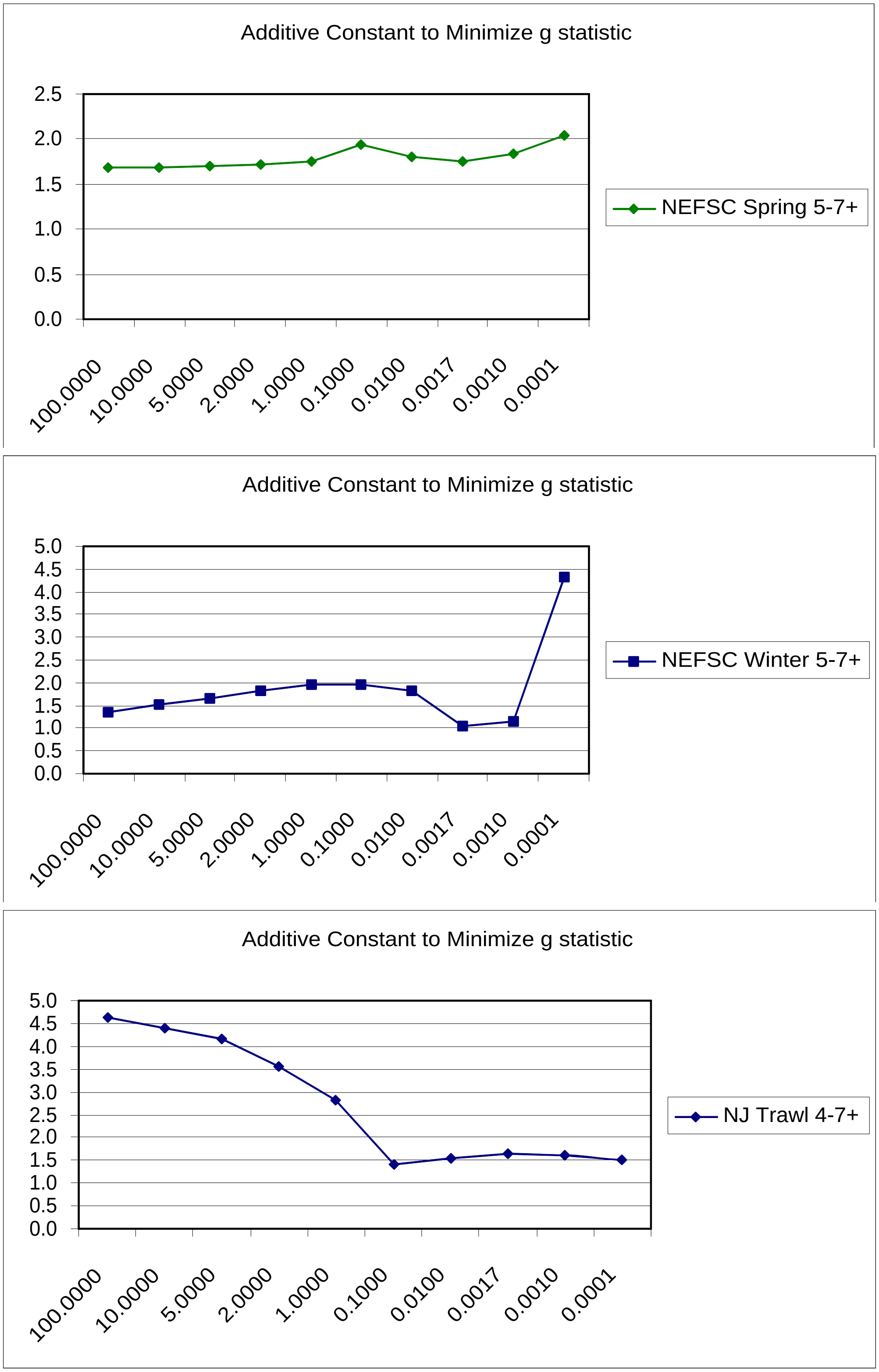

Additive Constant to Minimize g statistic
2.5
2.0
1.5
1.0
0.5
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0017
0.0010
0.0001
NEFSC Spring 5-7+
Additive Constant to Minimize g statistic
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0017
0.0010
0.0001
NEFSC Winter 5-7+
Additive Constant to Minimize g statistic
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
100.0000
10.0000
5.0000
2.0000
1.0000
0.1000
0.0100
0.0017
0.0010
0.0001
NJ Trawl 4-7+

Figure 6.