Icelandic saithe: New model to predict current weight at age

Arni Magnusson

1 May 2012

Abstract

A new regression model (Model 8) for predicting the current year's weight at age is presented. It overcomes some of the shortcomings of the previous model (Model 5), and decreases the potential prediction error for the younger ages, where most of the biomass tends to be. Overall, the choice between regression models does not make a big difference in the final biomass estimation.

Contents

1 Introduction

We have estimated numbers at age for the current year, and we multiply them with weight at age to compute the estimated biomass (Table 1).

Table 1. Estimated numbers, weights, and biomass from 2012 saithe assessment.

	3	4	5	6	7	8	9	10	11	12	13	14
N_{2012}	25092	48101	22577	14751	5265	1927	2465	2586	517	831	425	206
w_{2012}	1291	1629	2390	3298	4623	5680	6570	6276	6986	7881	9170	10026
B_{2012}	32393	78357	53959	48650	24338	10943	16196	16229	3609	6548	3898	2068

In the Icelandic saithe assessment, weight at age is based on samples of landed catch, and we don't have the 2012 landed catch, yet. Therefore we predict the current year's weight at age, using a regression model.

Since the predicted weights have a direct effect on the estimated current biomass, it is important to use a regression model that is relatively accurate for the most important ages (ages 4 to 9), and the worst-case performance of the model should also be considered. Model performance can be evaluated retrospectively, by comparing actual observations to what a given model would have predicted at the time.

A simpler approach is used to predict the current mean weight of the youngest and oldest individuals, ages 3 and 10–14, by taking the average weight of the three most recent years.

2 Methods

2.1 Historical catch weights

Table 2. Observed mean weight at age in historical catches.

	3	4	5	6	7	8	9	10	11	12	13	14
1980	1428	1983	2667	3689	5409	6321	7213	8565	9147	9617	10066	11041
1981	1585	2037	2696	3525	4541	6247	6991	8202	9537	9089	9351	10225
1982	1547	2194	3015	3183	5114	6202	7256	7922	8924	10134	9447	10535
1983	1530	2221	3171	4270	4107	5984	7565	8673	8801	9039	11138	9818
1984	1653	2432	3330	4681	5466	4973	7407	8179	8770	8831	11010	11127
1985	1609	2172	3169	3922	4697	6411	6492	8346	9401	10335	11027	10644
1986	1450	2190	2959	4402	5488	6406	7570	6487	9616	10462	11747	11902
1987	1516	1715	2670	3839	5081	6185	7330	8025	7974	9615	12246	11656
1988	1261	2017	2513	3476	4719	5932	7523	8439	8748	9559	10824	14099
1989	1403	2021	2194	3047	4505	5889	7172	8852	10170	10392	12522	11923
1990	1647	1983	2566	3021	4077	5744	7038	7564	8854	10645	11674	11431
1991	1224	1939	2432	3160	3634	4967	6629	7704	9061	9117	10922	11342
1992	1269	1909	2578	3288	4150	4865	6168	7926	8349	9029	11574	9466
1993	1381	2143	2742	3636	4398	5421	5319	7006	8070	10048	9106	11591
1994	1444	1836	2649	3512	4906	5539	6818	6374	8341	9770	10528	11257
1995	1370	1977	2769	3722	4621	5854	6416	7356	6815	8312	9119	11910
1996	1229	1755	2670	3802	4902	5681	7182	7734	9256	8322	10501	11894
1997	1325	1936	2409	3906	5032	6171	7202	7883	8856	9649	9621	10877
1998	1347	1972	2943	3419	4850	5962	6933	7781	8695	9564	10164	10379
1999	1279	2106	2752	3497	3831	5819	7072	8078	8865	10550	10823	11300
2000	1367	1929	2751	3274	4171	4447	6790	8216	9369	9817	10932	12204
2001	1280	1882	2599	3697	4420	5538	5639	7985	9059	9942	10632	10988
2002	1308	1946	2569	3266	4872	5365	6830	7067	9240	9659	10088	11632
2003	1310	1908	2545	3336	4069	5792	7156	8131	8051	10186	10948	11780
2004	1467	1847	2181	2918	4017	5135	7125	7732	8420	8927	10420	10622
2005	1287	1888	2307	2619	3516	5080	6060	8052	8292	8342	8567	10256
2006	1164	1722	2369	2808	3235	4361	6007	7166	8459	9324	9902	9636
2007	1140	1578	2122	2719	3495	4114	5402	6995	7792	9331	9970	10738
2008	1306	1805	2295	2749	3515	4530	5132	6394	7694	9170	9594	11258
2009	1412	1862	2561	3023	3676	4596	5651	6074	7356	8608	9812	10639
2010	1287	1787	2579	3469	4135	4850	5558	6289	6750	7997	9429	10481
2011	1175	1801	2526	3680	4613	5367	5685	6466	6851	7039	8268	8958

2.2 Regression data

The data used in the regression models are shown in Table 3, where `cW` is the observed mean weight at age in the catch from a given year, `cWlast` is the observed mean weight of the same cohort in the previous year's catch, and `sW` is the observed mean weight in the survey.

Table 3. Data used to fit the regression models.

Year	Age	cW	cWlast	sW
1985	4	2172	1653	1675
1985	5	3169	2432	2149
1985	6	3922	3330	3134
...				
2012	9	NA	5367	6769

The `cW` in the current year has not been observed yet, and are therefore shown as `NA` values. The objective is to predict `cW` in the current year for ages 4 to 9.

2.3 Model 5 (old model)

The regression model used in recent years is known by the name Model 5 from Sigurdsson and Steinarsson (2010),

$$ {}_{C}W_{t,a} \;=\; \beta_0 \;+\; \beta_1 \times {}_{C}W_{t-1,a-1} \;+\; \beta_2 \times {}_{S}W_{t,a} \;+\; \alpha_t \tag{1} $$

where β_0 is an overall intercept, β_1 and β_2 are slope coefficients, and α_t are individual "year effects", reflecting environmental conditions varying between years. This model is fitted using the following R command:

```
model5 <- lm(cW ~ cWlast + sW + factor(Year))
```

When using Model 5 to predict the current year's weight at age, it is assumed that this year's year effect is similar to last year's.

2.4 Model 8 (new model)

The proposed model, dubbed Model 8, log transforms the response and predictors, and does not have year effects,

$$ \log({}_{C}W_{t,a}) \;=\; \beta_0 \;+\; \beta_1 \times \log({}_{C}W_{t-1,a-1}) \;+\; \beta_2 \times \log({}_{S}W_{t,a}) \tag{2} $$

resulting in a model with only three regression parameters: an intercept and two slope coefficients. This model is fitted using the following R command:

```
model8 <- lm(log(cW) ~ log(cWlast) + log(sW))
```

3 Results

3.1 Response and predictors

The response `cW` shows a close relationship with both predictors `cWlast` and `sW` (Figure 1). The relationship between `cW` and `sW` (upper right panel) shows increasing variance at higher values, but the variance is more constant after log transformation (lower right panel).

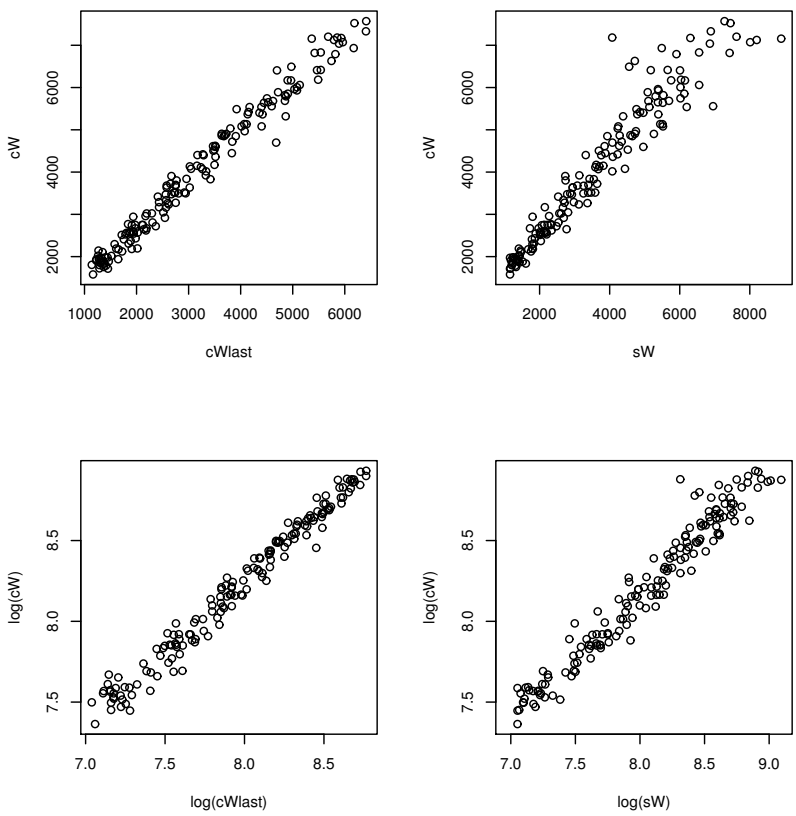

Figure 1. Relationship between the response `cW` and the predictors `cWlast` and `sW`, shown on a linear (upper panel) and log scale (lower panel).

3.2 Prediction performance

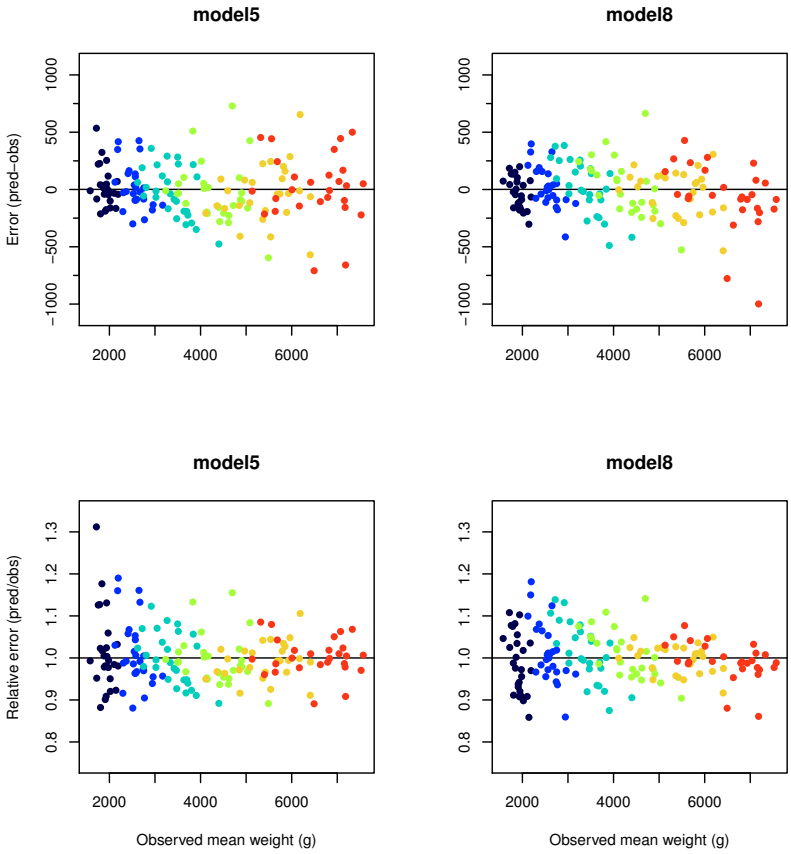

Figure 2. Prediction errors from the two regression models, in grams (upper panel) and on a relative scale (lower panel). Colors indicate ages 4 to 9.

Table 4. Standard deviation of prediction errors (pred–obs) in grams.

	4	5	6	7	8	9
Model 5	171	194	220	267	250	294
Model 8	137	175	234	248	193	300

Table 5. Standard deviation of relative prediction errors (pred/obs).

	4	5	6	7	8	9
Model 5	0.088	0.076	0.063	0.057	0.044	0.046
Model 8	0.072	0.069	0.069	0.055	0.035	0.047

3.3 Year effects

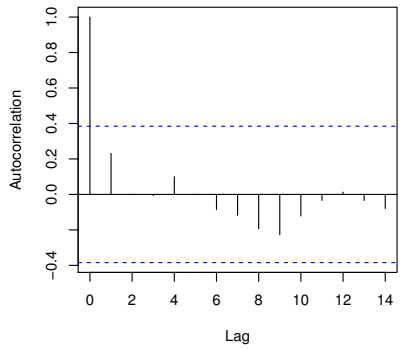

Figure 3. Autocorrelation of year effects estimated in Model 5.

The year effects estimated in Model 5 range from −174 to +294. They are not significantly autocorrelated (Box-Pierce test, $p = 0.24$).

3.4 Effect on biomass estimation

Table 6. Retrospective estimation of B4+ using Models 5 and 8 to predict the current year's weight at age.

	B4+ using historical weights	B4+ using Model 5 weights	B4+ using Model 8 weights	Model 5 minus historical	Model 8 minus historical
1985	299	300	300	1	0
1986	318	302	303	-17	-15
1987	335	382	348	47	13
1988	416	391	414	-25	-2
1989	398	417	415	19	17
1990	378	379	379	1	1
1991	336	350	350	14	14
1992	288	271	280	-17	-8
1993	231	227	221	-3	-9
1994	187	199	193	12	6
1995	153	147	149	-6	-3
1996	149	151	148	2	0
1997	156	156	146	1	-9
1998	153	152	141	-1	-12
1999	131	139	134	7	3
2000	142	138	142	-4	1
2001	161	156	161	-5	0
2002	216	222	213	6	-3
2003	274	274	269	-1	-5
2004	315	339	335	24	20
2005	279	279	298	0	20
2006	301	295	316	-6	15
2007	267	267	282	0	15
2008	234	225	234	-9	0
2009	211	209	210	-2	-1
2010	219	225	222	6	3
2011	234	231	230	-3	-4

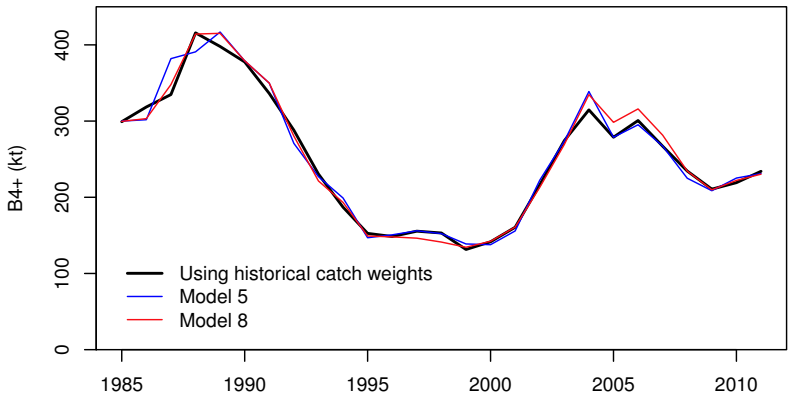

Figure 4. Retrospective estimation of B4+ using Models 5 and 8 to predict the current year's weight at age.

A retrospective comparison (Table 6, Figure 4) of using Models 5 and 8 vs. the weights that were eventually observed shows that Model 5 has greater deviations overall (standard deviation of residuals 13.8 kt) than Model 8 (standard deviation of residuals 9.8 kt). The relative deviations are also greater in Model 5 than in Model 8 (standard deviation of log residuals 0.045 vs. 0.037).

4 Discussion

The purpose of the new regression model is to overcome the following shortcomings of the old model:

- When fitting Model 5, it is equally influenced by a 500 gram residual, regardless of whether it is a 4 year old (2 kg) or 9 year old (6 kg) individual. Therefore, Model 5 predictions tend to have a large relative error for the younger ages (Figure 2 lower left panel, Tables 4 and 5).
- Model 5 assumes that the current year's "year effect" is similar to last year's effect, which can lead to very wrong predictions. For example, the +500 gram prediction error of 4-year-olds seen in Figure 2 (lower right panel) is mainly due to a highly positive year effect (+294) in one year, and a highly negative year effect (−153) in the next year, resulting in a large relative error for the younger ages.
- Year effects estimated in Model 5 do not indicate autocorrelation, so assuming that the current year's "year effect" is similar to last year's effect is not helpful when predicting the current year's weights.

Using only three regression parameters, Model 8 leads to slightly less deviations for the purposes of estimating the current biomass (Table 6). Compared to Model 5, the design of Model 8 decreases the potential prediction error for the younger ages, where most of the biomass tends to be. Overall, the choice between regression models does not make a big difference in the final biomass estimation (Figure 4).

5 References

Jonsson, S.T. and B.A. Steinarsson. 2010. Predicting current or assessment year weight at age in iSaithe landings. ICES NWWG WD 19.