King Salmon Sizes

Ben Buzzee

April 4, 2019

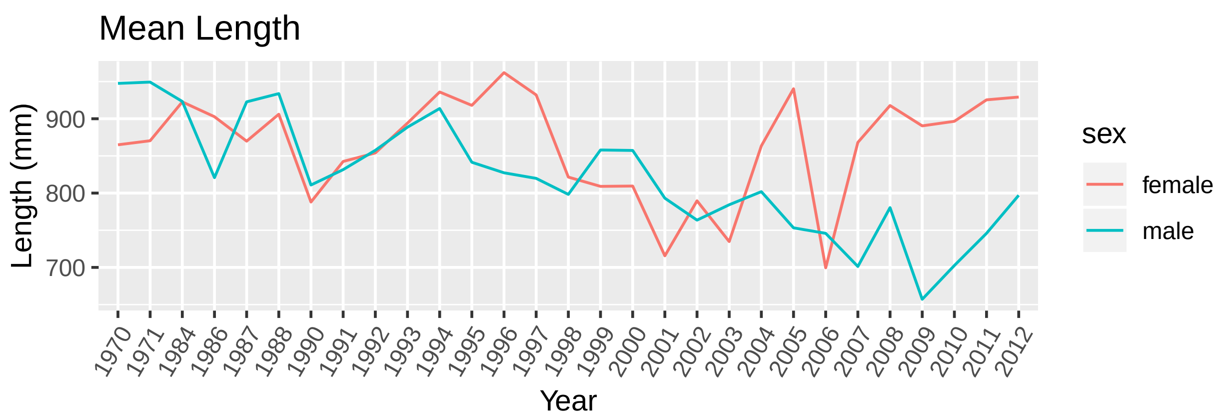
Background

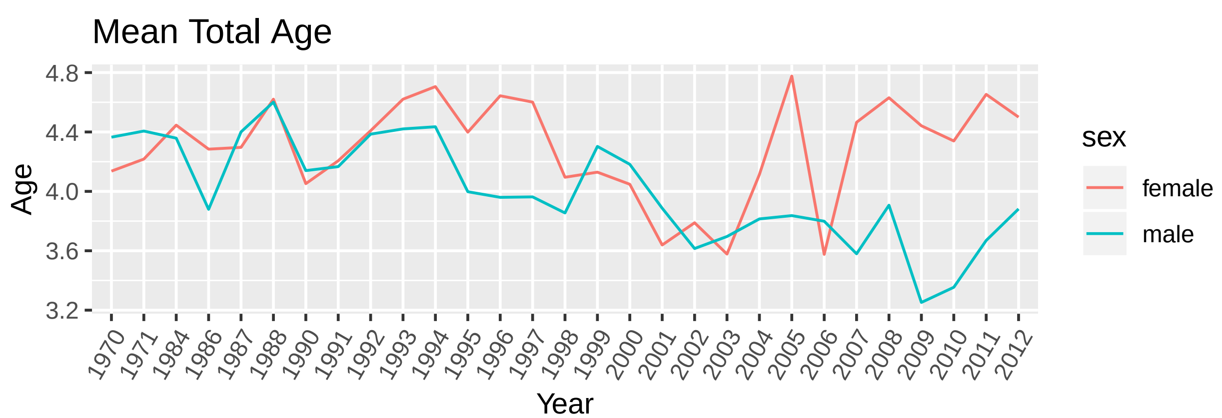
Chinook salmon are a pillar of the Cook Inlet sport fishery. As someone who is new to the area, I've heard a lot about the decline in size and numbers of the species. We will look at Chinook salmon age, sex, and length data collected from 1970 to 2012 in eastern Cook Inlet (District 244) by the commercial fish division of the ADFG to see if there are any trends and if forecasting the mean length of future returns is possible.

The data and documentation can be found at:https://knb.ecoinformatics.org/view/doi:10.5063/F18K77BT

Plotting and Data Exploration

As a first step in our time series analysis we will take a look at the raw data.



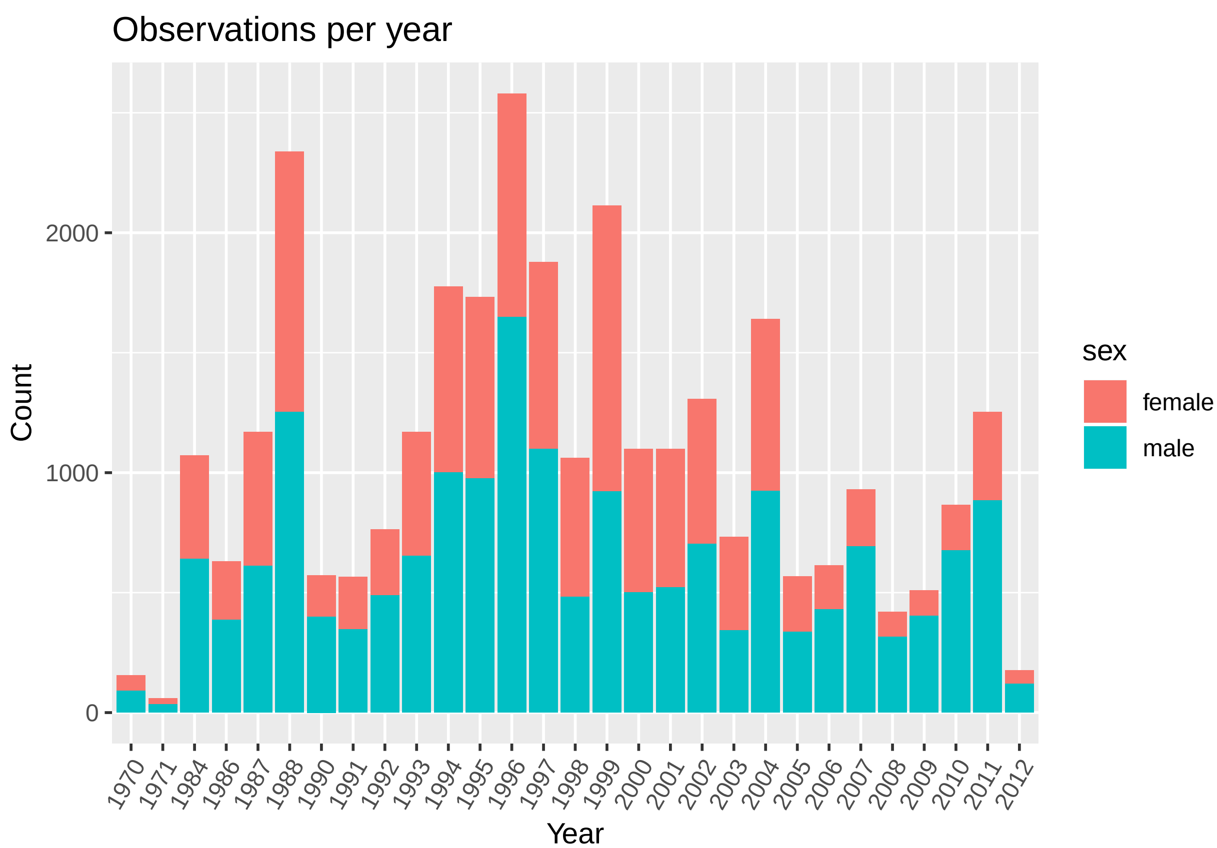


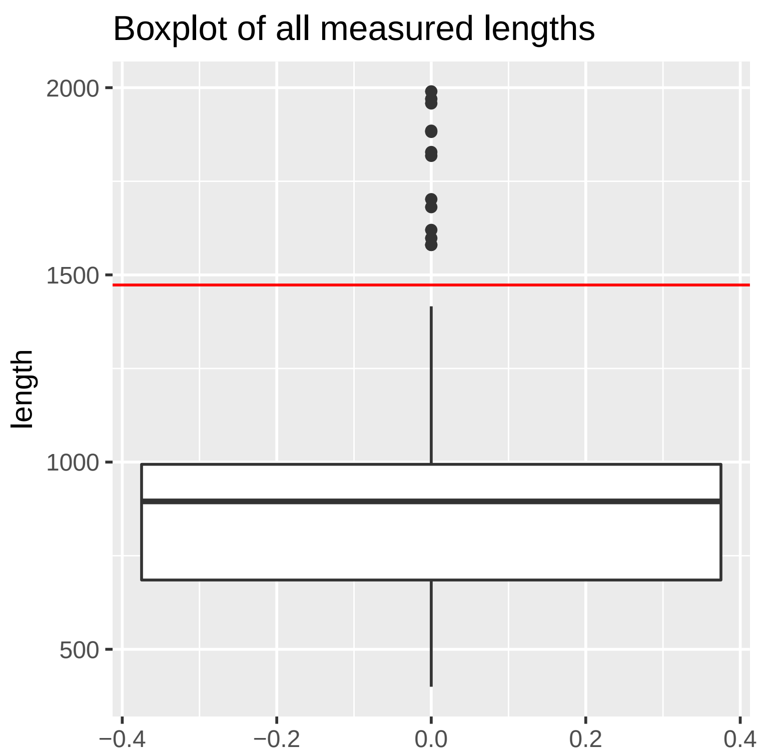
Comments on plots

By visually inspecting the data we can see that there seems to be a very strong correlation between length and age, which is to be expected. We also see that male and female fish seem to be following different trends. Both males and females seemed to decline in age and size in the late 90s and early 2000s. Since then, females have recovered back to their long run average, and males have continued to decline.

The decline in size and age of males appears to be quite significant. Compared to the 1980s, male chinook salmon appear to be 6 months to 1 year younger and 100mm - 200mm smaller, on average.

Data Quality Check





The current world record caught on the Kenai in 1985 was 58 inches, represented by the red line above. 2000 mm is roughly 6.5 feet. I'm extremely suspicious of the observations above the line. But all of the 1500+mm fish were supposedly caught between 1984 and 1988 and don't affect the trend over the last couple decades.

Research Questions:

Is the apparent decline in sizes of males statistically significant? Could it reasonably be attributed to noise?

Can we take advantage of any lagged linear relationships to accurately forecast the direction of future declines?

Stationarity and Trends

First we will check the length time series' for males and females to see if the time series are stationary and if there is any statistically significant trend.

ADF Test for Males:

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -100.393  -16.841    1.361   20.708   71.392
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1273.4026   286.0569   4.452 0.000183 ***
## z.lag.1       -1.3724     0.3020  -4.545 0.000145 ***
## tt           -10.2446     2.7010  -3.793 0.000939 ***
## z.diff.lag     0.3893     0.2115   1.841 0.078522 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.03 on 23 degrees of freedom
## Multiple R-squared:  0.5295, Adjusted R-squared:  0.4681
## F-statistic: 8.627 on 3 and 23 DF,  p-value: 0.0005103
##
##
## Value of test-statistic is: -4.5448 7.5553 10.8698
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -4.15 -3.50 -3.18
## phi2  7.02  5.13  4.31
## phi3  9.31  6.73  5.61
```

With a test statistic of -4.54, we can reject the null hypothesis that male chinook lengths are characterized by a random walk. And with a test statistic of 10.87, we also find statistically signficant evidence of a time trend.

ADF Test for Females:

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -190.14  -28.21   17.91   47.40   88.05
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 503.07845  223.03265   2.256   0.0339 *
## z.lag.1      -0.57992    0.25017  -2.318   0.0297 *
## tt           -0.04378    1.80134  -0.024   0.9808
## z.diff.lag   -0.09778    0.21571  -0.453   0.6546
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

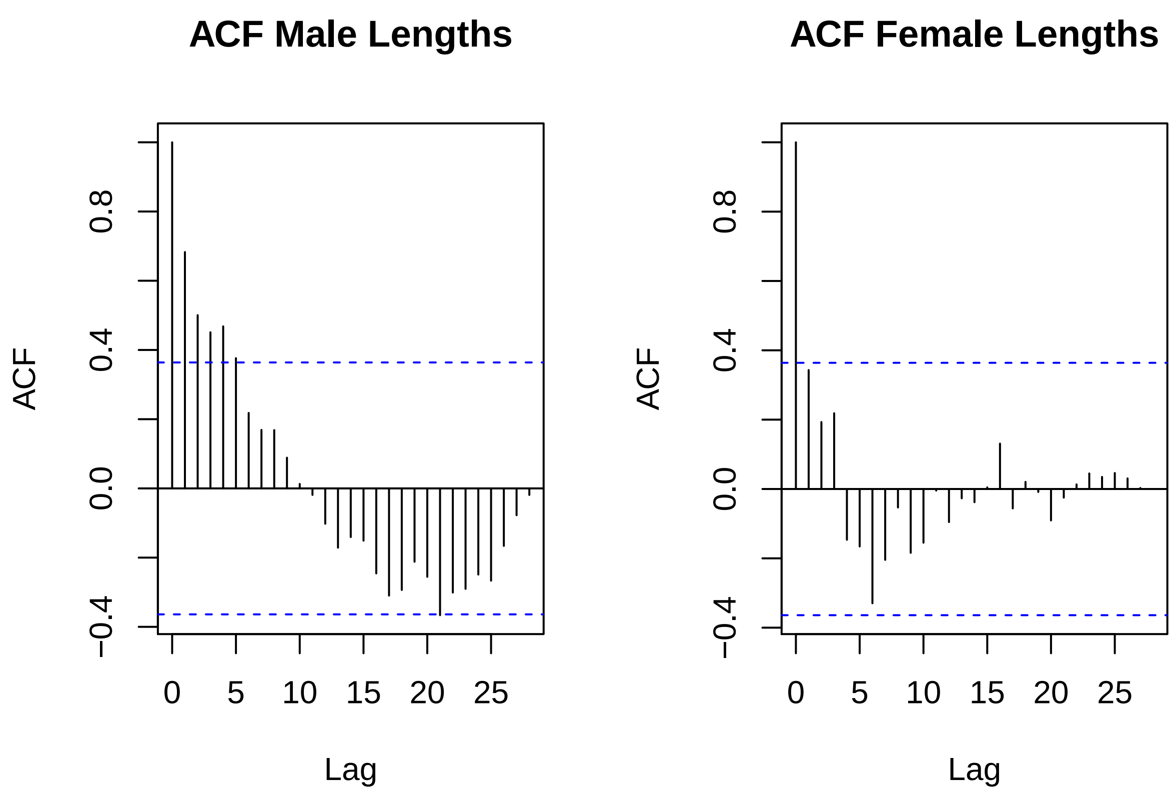
```
## Residual standard error: 71.2 on 23 degrees of freedom
## Multiple R-squared:  0.3211, Adjusted R-squared:  0.2325
## F-statistic: 3.626 on 3 and 23 DF,  p-value: 0.0281
##
##
## Value of test-statistic is: -2.3181 1.8811 2.7967
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau3 -4.15 -3.50 -3.18
## phi2  7.02  5.13  4.31
## phi3  9.31  6.73  5.61
```

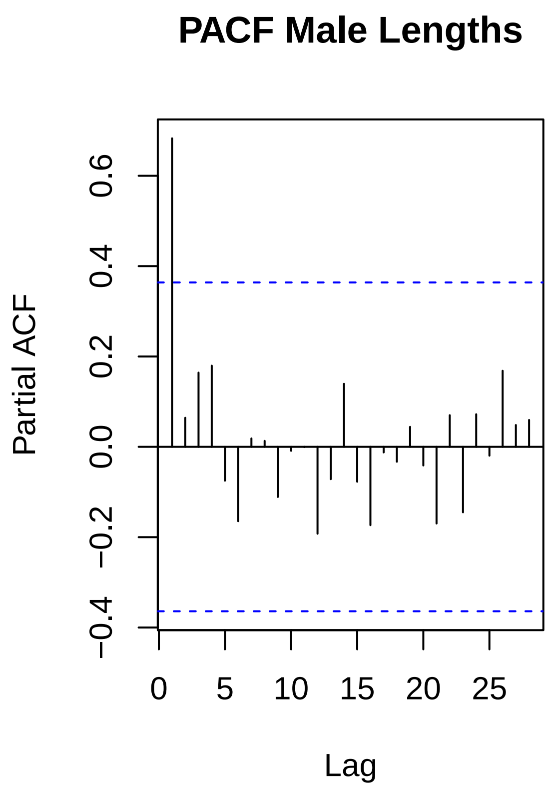
With a test statistic of -2.318 and p-value of .0297, we also find that female chinook lengths are not characterized by a random walk.

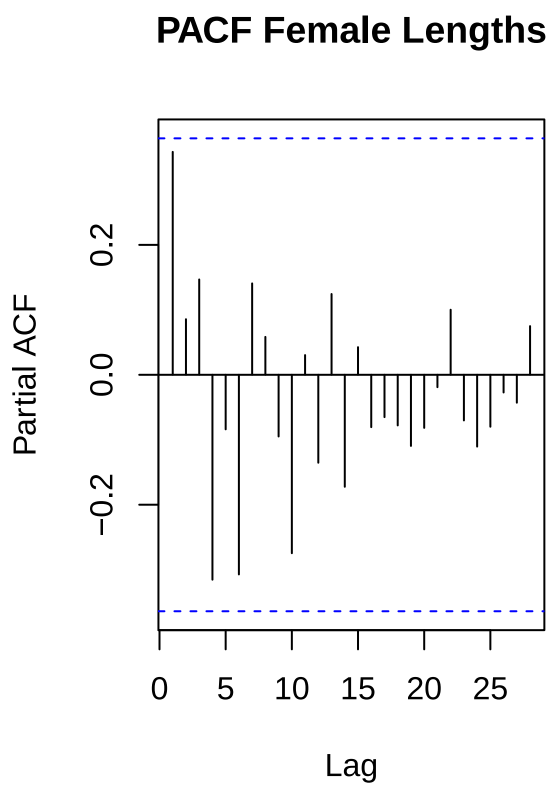
Model Order

To explore possible lagged autoregressive relationships, we will take a look at ACF and PACF plots.

ACF and PACF Plots







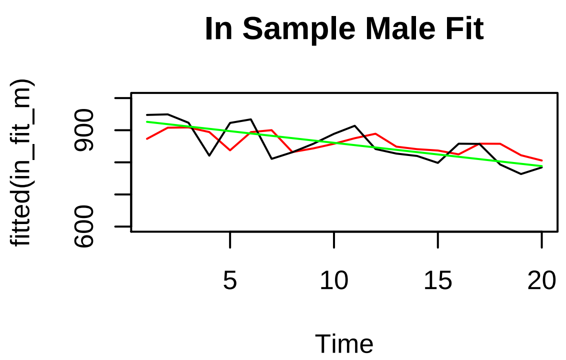
Interpretation of ACF and PACF plots

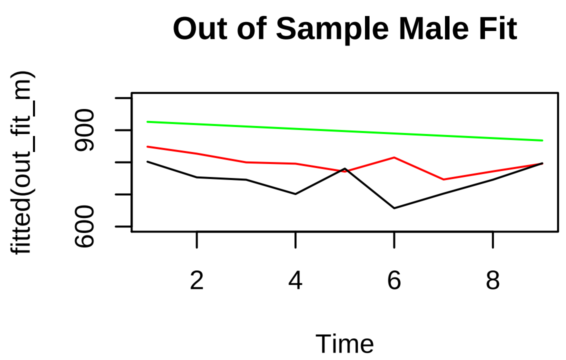
For males, we see the ACF plot tailing off, and the PACF abrubtly cut off after a lag of 1. This may be a sign that an AR(1) model could be a good model for this time series. For female Chinook, there appears to be no discernable lagged relationship.

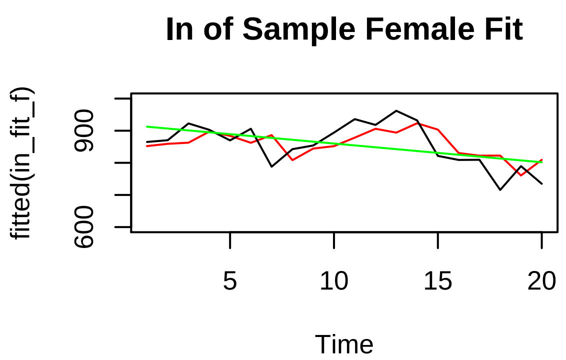
Simple AR Model

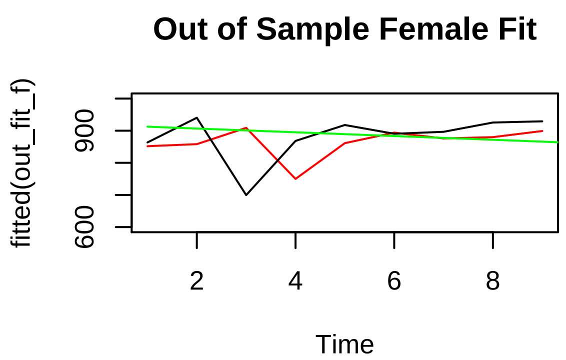
Next we will check the ability of a simple model to forecast future mean lengths. We will fit AR(1) models to both male and female lengths and compare to a simple time trend model. We will judge each model by out-of-sample MSE.

Note: Red is the one-year-ahead AR(1) forcasted mean length, black is the observed mean length, green is a linear model with time as the explanatory variable.









MSE & Testing for Model Deterioration

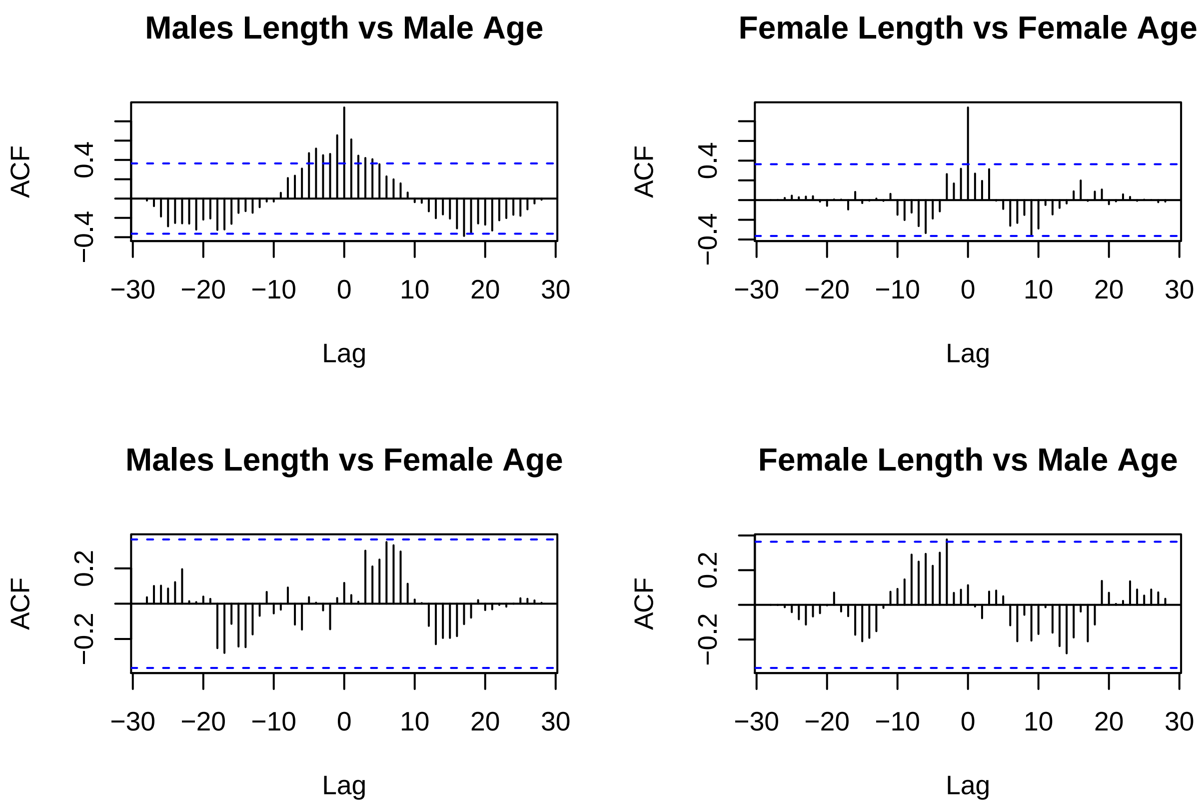
model	in_sample_mse	out_of_sample_mse	f_stats	p_vals
Time Trend Males	1434.753	25930.78	18.07	0.00
Time Trend Females	3244.762	6393.92	1.97	0.10
AR(1) Males	2308.595	5231.25	2.27	0.06
AR(1) Female	2604.003	7925.10	3.04	0.02

After conducting and F-Test for model deterioration, we see that all the models deteriorated somewhat. Using a significane level of of .05, we would fail to reject the null that the in-sample and out-of-sample MSEs are different for the linear time trend model for female chinook and the AR(1) model for male chinook.

Multivariate Analysis

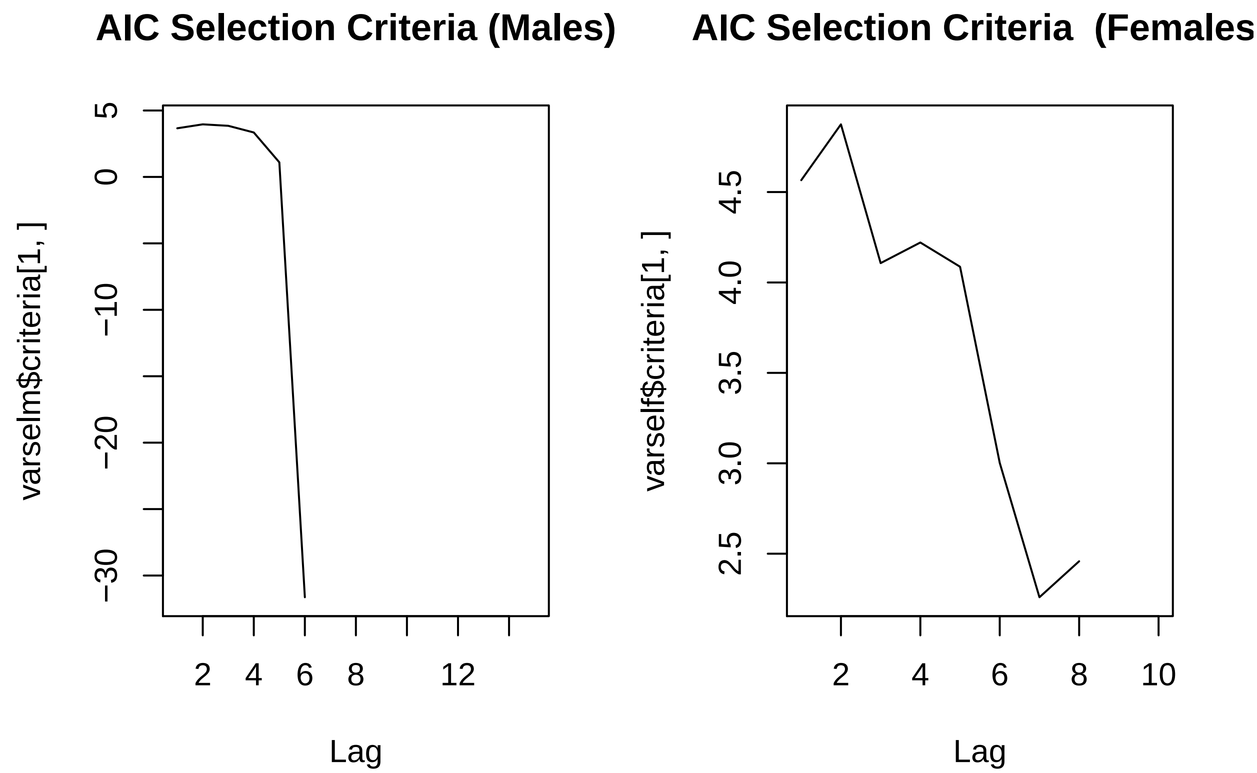
Next we will see if we can improve our forecast by incorporating any lagged linear relationships among age and length. To explore this possibility we will look at CCF plots.

CCF Plots



We see from the CCF plot there is a clear zero-lag correlation between age and length within sex. There may also be some ability to forecast future male lengths using previous values of their age.

Based on the above plots, I do not expect for there to be any lagged relationship between age and length. We can check what would happen if we tried to fit a model anyway:



Both selection criteria shoot off to -Infinity at lag 9. This suggests a problem with calaculating the AIC (the same occurs for other the selection criteria) and is not indicative of a selected model order. We can try fitting a simple VAR model just to see what happens. First for male chinook with length as the response:

```
##
## Call:
## lm(formula = y ~ -1 + ., data = datamat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -126.275  -33.840    0.101   36.980   99.179
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## mean_len.l1   0.9531     0.3921   2.431   0.0226 *
## mean_age.l1 -62.8843    87.1382  -0.722   0.4772
## const       285.2172   116.5309   2.448   0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.51 on 25 degrees of freedom
## Multiple R-squared:  0.5312, Adjusted R-squared:  0.4937
## F-statistic: 14.16 on 2 and 25 DF,  p-value: 7.72e-05
```

We see that only a constant and a lag of length is signficant, which is just an AR(1) model.

Next we can try the same for females with length as the response:

```
##
## Call:
## lm(formula = y ~ -1 + ., data = datamat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -126.275  -33.840    0.101   36.980   99.179
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## mean_len.l1   0.9531     0.3921   2.431   0.0226 *
## mean_age.l1 -62.8843    87.1382  -0.722   0.4772
## const       285.2172   116.5309   2.448   0.0217 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 52.51 on 25 degrees of freedom
## Multiple R-squared:  0.5312, Adjusted R-squared:  0.4937
## F-statistic: 14.16 on 2 and 25 DF,  p-value: 7.72e-05
```

Again, there is no lagged relationship with age.

Conclusion

From our ADF test, we can conclude that male chinook lengths are not characterized by a random walk. Further, we found evidence of a statistically significant downward trend. We found that a simple AR(1) model provided improved forecasts over a simple time trend model. In exploring possible lagged relationships between length and age, we did not find any.

Likewise for female chinook we found their length time series was not a random walk. For their time series, however, a simple time trend offered superior forecasting ability over an AR(1) model. For female chinook as well, there was no meaninful lagged relationship between age and length.

As a next step we could look into incorporate environement variables in our analysis to improve forecastability and explore possible causal relationships.