Model selection

We'll start just fitting a model with just an intercept, and then add in effects of gears,

```
null.model = glm(SEX ~ 1, family = "binomial", data=d)
gear.model = glm(SEX ~ gear, family = "binomial", data=d)
```

We can add time in in one of two ways, either as a linear trend or a factor. [Note, we could also use non-linear models, like GAMs, but those give a pretty linear estimate because of small sample sizes]

```
gearTime.model01 = glm(SEX ~ gear + as.factor(Year), family = "binomial", data=d)
gearTime.model02 = glm(SEX ~ gear + Year, family = "binomial", data=d)
```

We can also try a few flavors of GAMs,

```
gam.model = gam(SEX ~ gear + s(Year), family = "binomial", data=d)
gam.model.02 = gam(SEX ~ s(Year, by=as.factor(gear)), family = "binomial", data=d)
```

Model	AIC
Null	170338.56
Gear	169236.57
Gear, time (factor)	168672.24
Gear, time (numeric)	169233.94
GAM s(year)	169087.22
GAM s(year, gear)	169024.76

Based on this table (lower AICs = better models) we'd say that the best model includes time as factor, but both that model and the model with the temporal trend are better than the model without time as a predictor. This model also does better than GAM models, with smooth terms (either shared, or by gear). To visualize what the sex ratio looks like, we can look at the output of a GAM.

Are these changes biologically meaningful? We can make predictions with a new data frame,

Correlation with pink salmon

Spatial comparison

One of the things we can do is look for matched sites in the same year and compare purse seine samples for them. After accounting for the variation over years and spatial sampling locations within PWS, the difference between gears is still significant.

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: SEX ~ gear + (1 | loc_year)
##    Data: d[which(d$loc_year %in% g$loc_year), ]
##
##      AIC      BIC   logLik deviance df.resid
##  36007.7  36032.3 -18000.9  36001.7    26706
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5091 -1.0758  0.7346  0.8749  1.2215
```

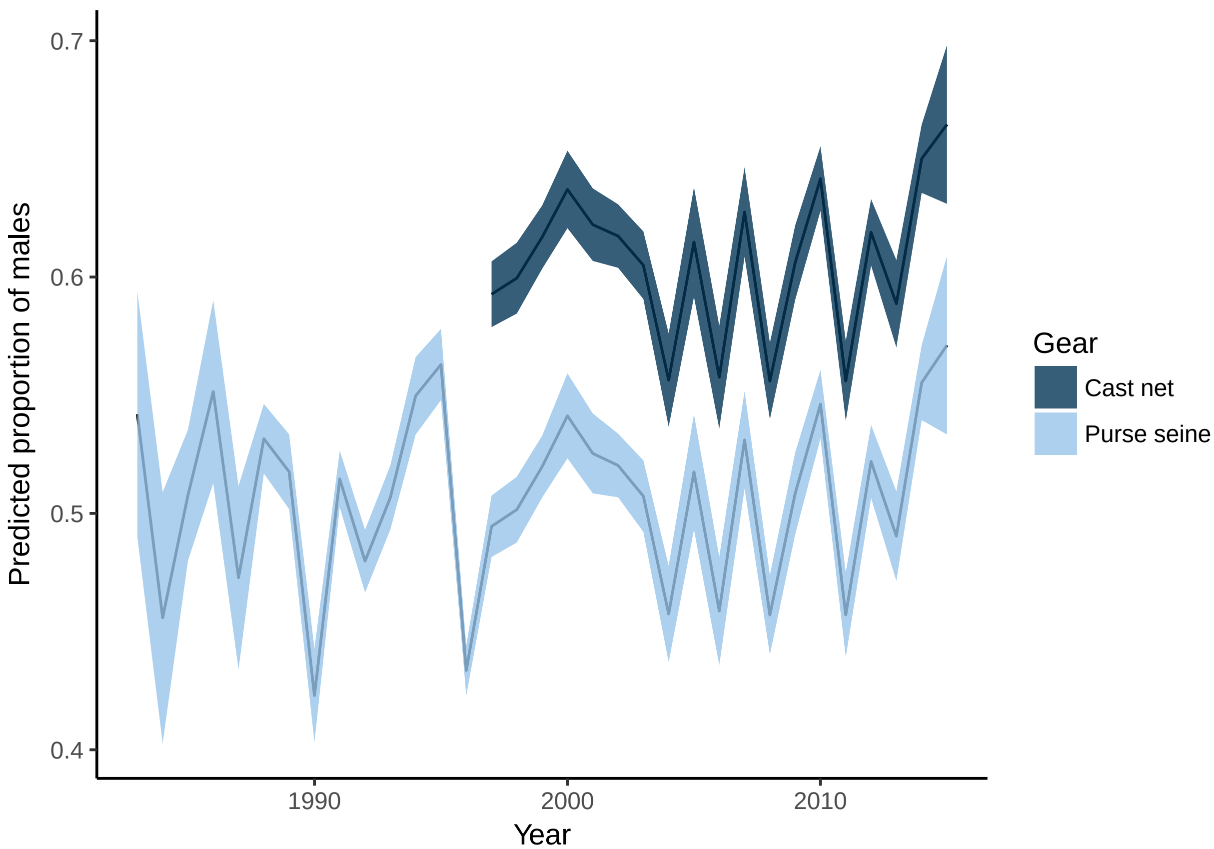

Figure 1: Estimated percent male for PWS herring, for 2 gears sampling in spring (March - April). Estimates shown from best-fit GLM (Table 1), where gear is treated as a factor/offset.

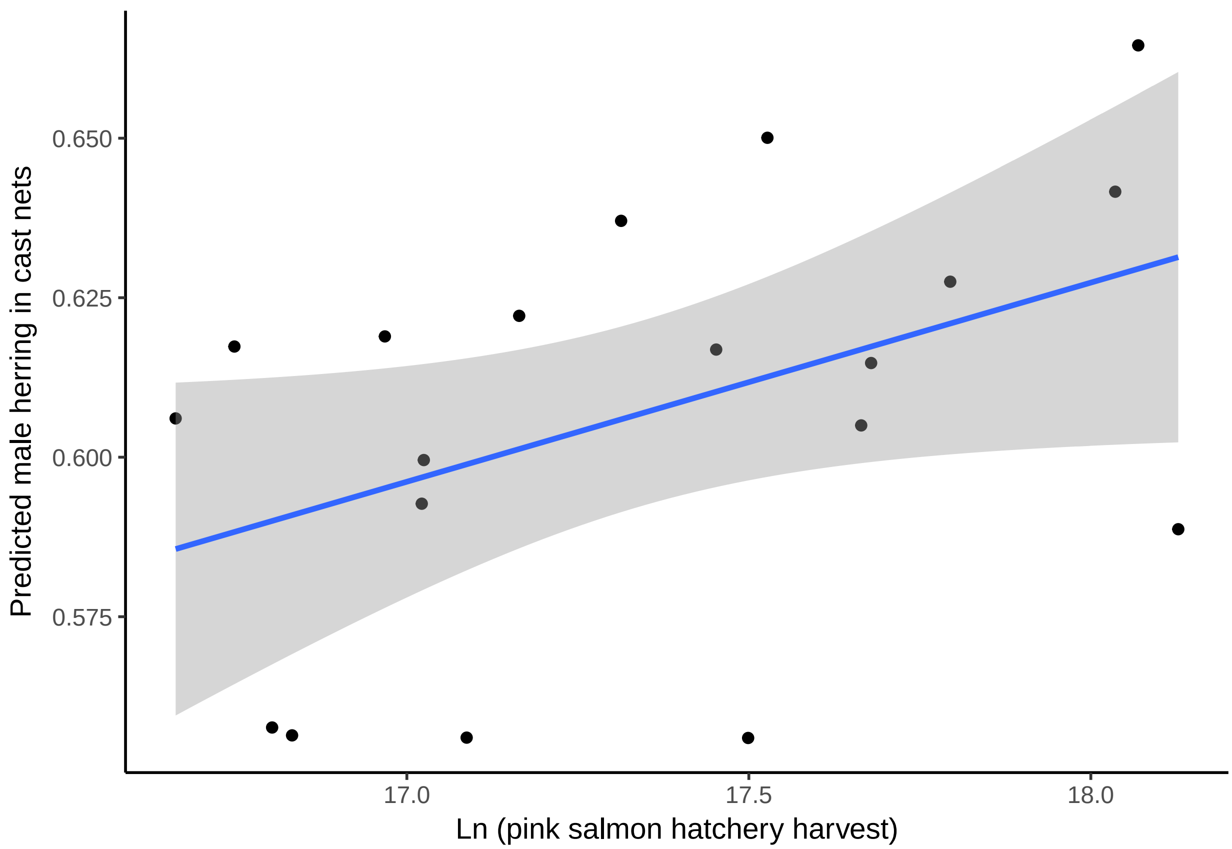

Figure 2: Correlation between ln(pink salmon harvest) and predicted male herring in cast nets

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  loc_year (Intercept) 0.04196  0.2048
## Number of obs: 26709, groups:  loc_year, 25
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.50585    0.04516   11.20   <2e-16 ***
## gearP       -0.46505    0.02633  -17.66   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##       (Intr)
## gearP -0.277
```

Year coefficients

The year coefficients from the best model show an interesting pattern of oscillations over time, mostly between odd / even years. Extracting these coefficients,

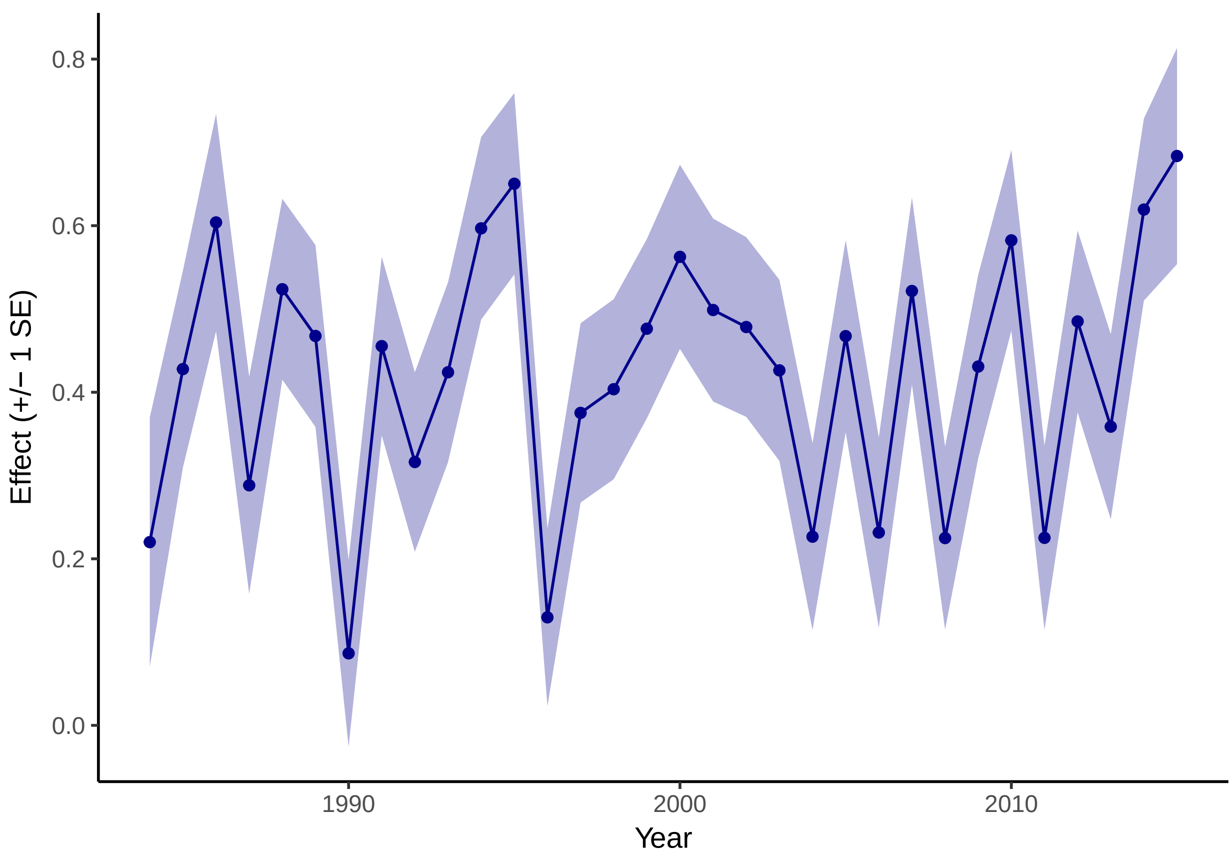

Figure 3: Estimated year coefficients from best-fit GLM (Table 1), where gear is treated as a factor/offset.

We can also look at these in the raw data, calculating proportions by age and gear.