Analysis notes

May 5, 2014

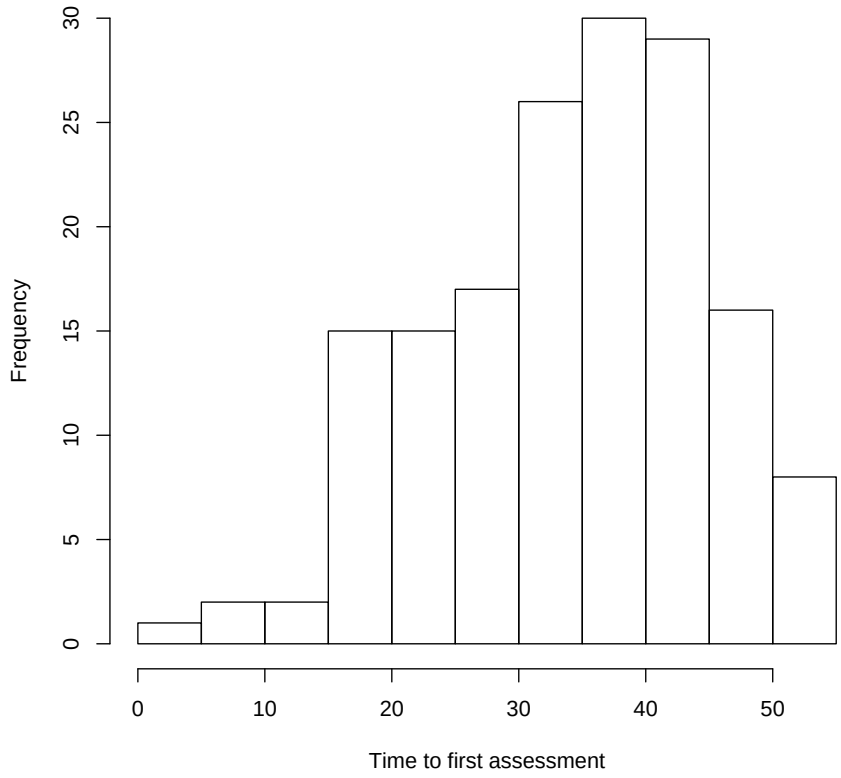
V0.0.1 PN

1 Data manupulation

Data were crudely transformed to input, using Mike's suggestions for fishery deveopment year.

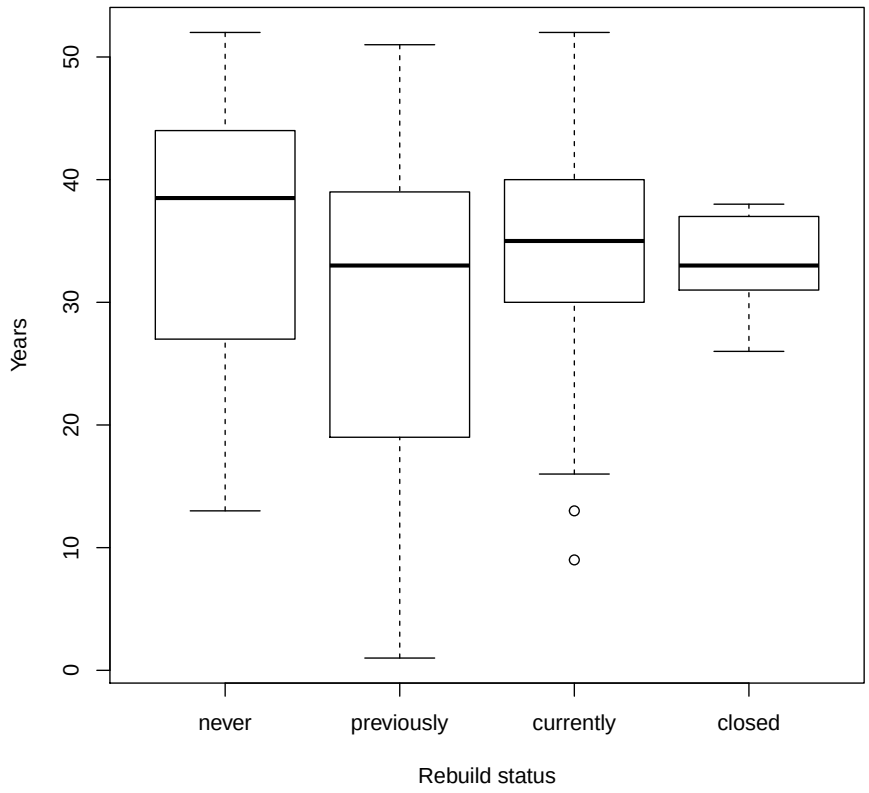
```
require(dplyr, quietly = T, warn.conflicts = F)

year.table <- tbl_df(read.csv("~/Work/Dropbox/First year of assessment/V3_Final_dataset.csv",
    na.strings = c("", "NA")))

ref.time <- min(year.table$Year.of.first.stock.assessment, na.rm = T) - 1

year.table <- year.table %.% mutate(time = Year.of.first.stock.assessment -
    ref.time, survey.abs = Year.of.first.fishery.indepen.dent.surveys)
```

```
hist(year.table$time, 10, main = "", xlab = "Time to first assessment")
```




```
year.table$Rebuild <- ordered(year.table$Rebuilding.plan.history, levels = c("never",
    "previously", "currently", "closed"))

with(year.table, table(Rebuild))

## Rebuild
##      never previously  currently     closed
##        115         20         51          7

with(year.table, table(Region))

## Region
## Alaska   USEC   USNE   USSE   USWC
##     51      9     49     34     50

# set north-east as reference treatment
RC <- contr.treatment(levels(year.table$Region), base = 3)
```

```
boxplot(time ~ Rebuild, data = year.table, xlab = "Rebuild status", ylab = "Years")
```



2 Try simple GLM, no censoring

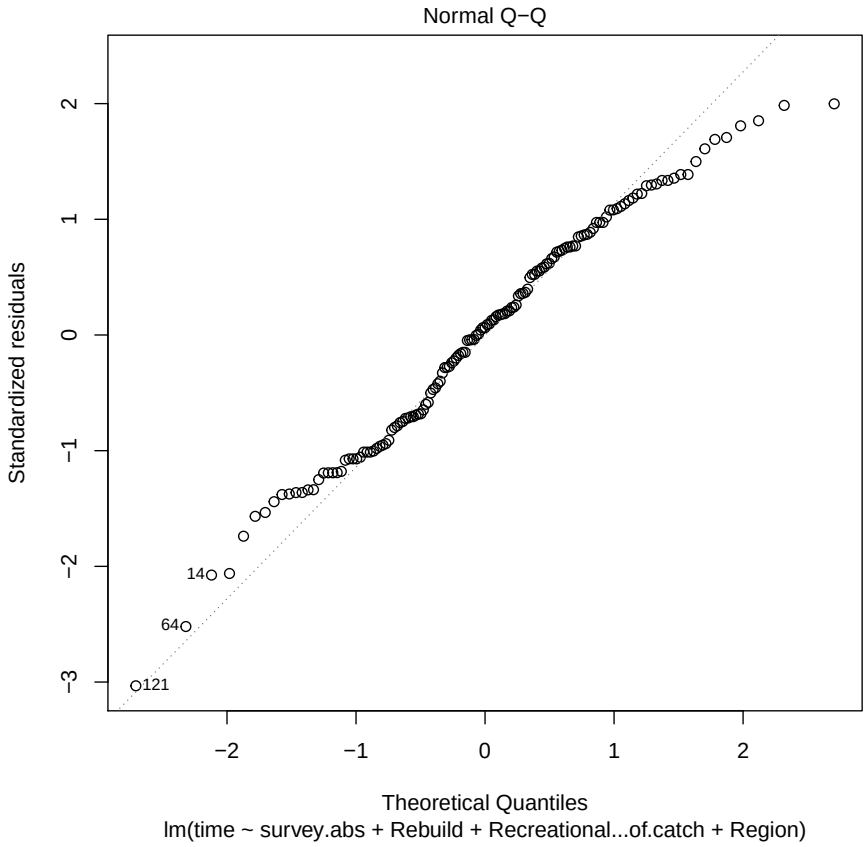
Try a simple normal linear model without any censoring

```
ylm.abs <- lm(time ~ survey.abs + Rebuild + Recreational...of.catch + Region,
    contrasts = list(Region = RC), data = year.table)

summary(ylm.abs)

##
## Call:
## lm(formula = time ~ survey.abs + Rebuild + Recreational...of.catch +
##     Region, data = year.table, contrasts = list(Region = RC))
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.147  -7.607   0.621   7.362  19.621
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             204.7571   197.6942    1.04    0.302
## survey.abs               -0.0894     0.1005   -0.89    0.375
## Rebuild.L                 0.0930     3.0107    0.03    0.975
## Rebuild.Q                 3.3488     2.8205    1.19    0.237
## Rebuild.C                -2.6146     2.0902   -1.25    0.213
## Recreational...of.catch   8.7174     4.1283    2.11    0.037 *
## RegionAlaska              4.3478     2.7646    1.57    0.118
## RegionUSEC               -0.9126     3.9399   -0.23    0.817
## RegionUSSE                6.0852     3.3876    1.80    0.075 .
## RegionUSWC                4.9579     2.8846    1.72    0.088 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10 on 137 degrees of freedom
##   (46 observations deleted due to missingness)
## Multiple R-squared:  0.133,Adjusted R-squared:  0.0762
## F-statistic: 2.34 on 9 and 137 DF,  p-value: 0.0175
```

The contrast in the status is with respect to the base case (never been in a rebuiling plan). The model doesn't explain much ($< 8\%$) and the fit is ok, but not great:

```
plot(ylm.abs, w = 2)
```



3 Simple Bayesian Weibull survival model

Try a Bayesian trunkated Weibull model to keep is simple to start with:

```
# assessment time
a.time <- year.table$time

# true false censoring
censored <- as.numeric(is.na(a.time))

# censor time - improve here from the arbitrary 2010 cutoff for censored
# (non-assessed stocks)
ctime <- a.time
ctime[is.na(a.time)] <- 2010 - ref.time
```

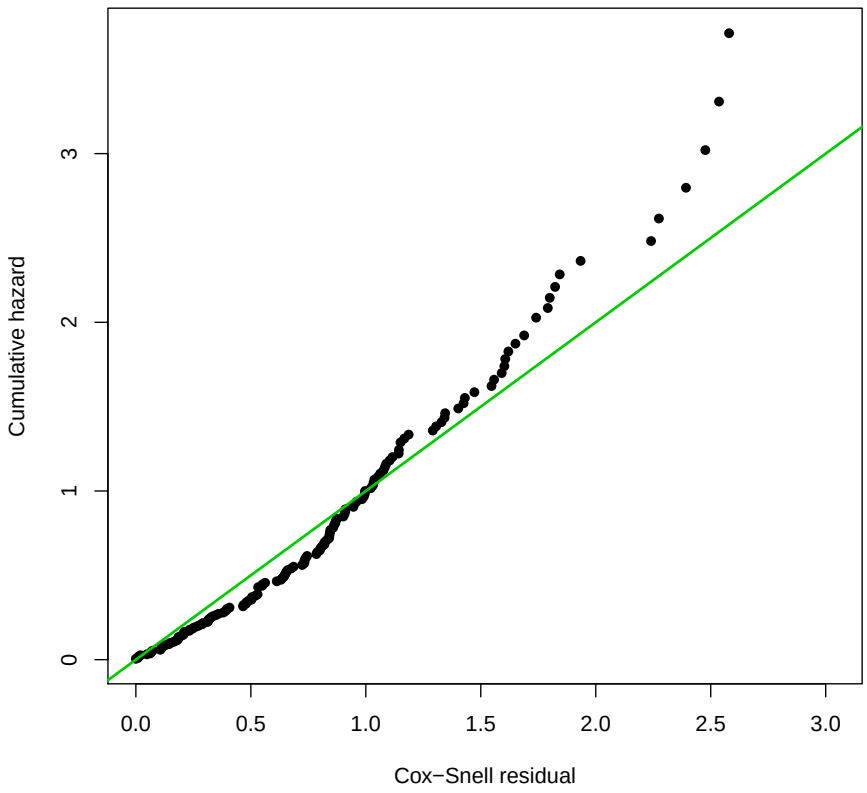
```r
# initial values for censored observations
time.inits <- ctime + 1
time.inits[!is.na(a.time)] <- NA

# habitat and family random effect - note - taxon is half way between
# habitat and family variables, try that too sometime

afs <- function(x) as.numeric(as.factor(x))

hab <- with(year.table, afs(habitat_MM))
n.hab <- length(unique(hab))

fam <- with(year.table, afs(Family))
n.fam <- length(unique(fam))

# fixed effect for regions
region <- data.frame(with(year.table, model.matrix(~Region, contrasts.arg = list(Region = RC
#

# Covariate dataframe
covs <- year.table %.% select(Lmax..cm., TL, Recreational...of.catch, Year.of.fishery.develo

# replace TL for CA spiny lobster with something approximate for now since I
# can't find a good value
covs$TL[is.na(covs$TL)] <- 3.2

# scale covariates for comparison
sc.covs <- data.frame(apply(covs, 2, function(x) (x - mean(x))/(2 * sd(x))))
COVS <- cbind(region, sc.covs)
n.covs <- ncol(COVS)
n.stocks <- nrow(COVS)
```

```r
# set up jags model

require(rjags)

jags.data <- list(COVS = COVS, n.covs = n.covs, n.stocks = n.stocks, hab = hab,
    fam = fam, n.hab = n.hab, n.fam = n.fam, ctime = ctime, a.time = a.time,
    censored = censored)

# run model - short run for now...
JM <- jags.model("Weib_surv.R", inits = list(a.time = time.inits), data = jags.data,
    n.chains = 1)

update(JM, n.iter = 2000)
```

```
a.out <- coda.samples(JM, variable.names = c("betas", "habitat", "fp.var", "CS"),
    n.iter = 10000, thin = 10)
```

```
plot(a.out)
crosscorr.plot(a.out)
```

```
# get coeffs from the chains - pull in some helper functions and Rdata from
# Bayesian model
source("helper_functs.R")
load("~/Work/Dropbox/First year of assessment/FA_V001.RData")

# get posterior for cox-snell(CS) residuals from MCMC
CS.full <- tbl_df(get_coef_chains(model.out = a.out, coef.names = "CS"))

# just look at mean CS for now, can put posterior around it later
CS.means <- CS.full %.% group_by(Parameter) %.% summarise(post.mean = mean(MCMC))

library(survival)

## Loading required package:  splines

CS = CS.means$post.mean

# Kaplan-Meyer non-parametric survival at CS - should follow exp(1)
# distribution
km.cs <- survfit(Surv(CS, !censored) ~ 1)
summary.km.cs <- summary(km.cs)
rcu <- summary.km.rc$time  # Cox-Snell residuals of
# uncensored points.
surv.cs <- summary.km.cs$surv
plot(rcu, -log(surv.cs), type = "p", pch = 16, xlab = "Cox-Snell residual",
    ylab = "Cumulative hazard")
abline(a = 0, b = 1, col = 3, lwd = 2)
```



It looks as though the fit of the Weibull isn't too bad, some deviation is expected in the tails of the distribution, but over the bulk it seems to follow the 1:1 line fairly closely. Perhaps not much of an improvement over the normal GLM above - should try differnet distributional assumptions here. We could also plot the posterior for the CS residuals and see if the 1:1 line falls outside a 95% interval for any points.

```
coeffs <- tbl_df(get_coef_chains(model.out = a.out, coef.names = "betas", var.names = colnam

# regressin coeffs are -beta
coef_P <- coeffs %.% group_by(Parameter) %.% summarise(post.mean = -mean(MCMC),
    post.P = 1 - mean(MCMC > 0))
```

```
# habitat
habs <- tbl_df(get_coef_chains(model.out = a.out, coef.names = "habitat", var.names = with(y
    levels(habitat_MM))))
```

Table 1: Posterior mean and $P(\beta > 0)$ for model parameters

Parameter	Posterior Mean	Bayesian P
Lmax..cm.	0.14	0.74
Recreational...of.catch	-0.17	0.23
RegionAlaska	0.35	0.83
RegionUSEC	-0.17	0.38
RegionUSSE	-0.06	0.46
RegionUSWC	0.80	0.99
TL	0.52	0.94
X.Intercept.	14.41	1.00
Year.of.fishery.development	0.26	0.86

Table 2: Posterior mean and $P(\beta > 0)$ for model habitat

Habitat	Posterior Mean	Bayesian P
benthic	0.20	0.66
benthopelagic	-0.11	0.28
demersal	0.20	0.68
pelagic	-0.10	0.35
reef	0.10	0.53

```
hab_P <- habs %.% group_by(Parameter) %.% summarise(post.mean = -mean(MCMC),
    post.P = 1 - mean(MCMC > 0))
```

```
# finite population variance of family random effects
fp.vars <- tbl_df(get_coef_chains(model.out = a.out, coef.names = "fp.var")) %.%
    summarise(mean(MCMC)^2)

fp.vars

## Source: local data frame [1 x 1]
##
##   mean(MCMC)^2
## 1       0.8259
```