SSHI sockeye exploratory analyses (stage 2)

2019-10-29

This document summarizes some quick exploratory analyses that relate Fraser sockeye survival indices to infectious agent data. All code and associated data can be found on Github here.

Infectious agent data provided by A. Teffer, Fraser sockeye data are described here. NA's in load (i.e., years with no infected individuals) replaced with 0.

For brevity I have turned off warnings. But they need to be carefully considered when digging into these models moving forward. See note below about great tool for understanding/visualizing MCMC samples from models fit using STAN.

load required packages and functions

```
library(lme4)
library(rstanarm) # https://mc-stan.org/users/documentation/case-studies/tutorial_rstanarm.html
library(ggplot2)
library(plotrix)
```

load data, standardize independent variables

```
inf_agt_resid_data <- read.csv("data/ONNE_productivity_infection_analysis.csv")
inf_std <- plyr::ddply(inf_agt_resid_data, c("agent"),function(x) {
                      scaled_prev <- scale(x$prev)
                      scaled_load <- scale(x$mean_load_all)
                      xx <- data.frame(scaled_prev, scaled_load)
    })

inf_agt_resid_data$prev_std <- inf_std[,2]
inf_agt_resid_data$load_std <- inf_std[,3]
agents <- unique(inf_agt_resid_data$agent)
inf_agt_resid_data$Stock <- inf_agt_resid_data$Stock_Analysis

head(inf_agt_resid_data)
```

```
##   Year orderID Stock_Analysis brood_year   metric resid_value  N N. prev
## 1 2009     546       L.Stuart       2007 SR_resid  -0.9099142 24  0    0
## 2 2009    1072         Weaver       2007 SR_resid   0.3732061 24  0    0
## 3 2009     844           Raft       2007 SR_resid  -1.6863897 24  0    0
## 4 2009    1023       Stellako       2007 SR_resid  -0.2379979 24  0    0
## 5 2009     416       Harrison       2007 SR_resid   1.3025948 24  0    0
## 6 2009     481      L.Shuswap       2007 SR_resid   0.3932027 24  0    0
##   mean_load prevload  agent mean_load_all   prev_std   load_std     Stock
## 1        NA       NA arena2             0 -0.6770613 -0.4500811  L.Stuart
## 2        NA       NA arena2             0 -0.6770613 -0.4500811    Weaver
## 3        NA       NA arena2             0 -0.6770613 -0.4500811      Raft
## 4        NA       NA arena2             0 -0.6770613 -0.4500811  Stellako
## 5        NA       NA arena2             0 -0.6770613 -0.4500811  Harrison
## 6        NA       NA arena2             0 -0.6770613 -0.4500811 L.Shuswap
```

A couple of key decisons to ponder before proceeding:

1. There are two alternatives modelling approaches we can try at this point. Both are generalized linear mixed effects modelling approaches, but one is fit using Maximum Likelihood methods (via lme4 R package) and the other is fit via Bayesian methods (via rstanarm R package). The formula syntax

used in both is nearly identical which makes their use very compliemntary. I illustrte both below but suggest going the rstanarm route because parameter estimates can be more intuitively interpreted probabilistically and because it allows for more complete estimation of random effects and uncertainty. Plus there are some great tools developed for model evaluation (e.g., see STAN shiny app here)

2. In addition to the decison about estimation framework, we also need to decide on the most sensible random effects structure. Initial thinking was that we want the effect of infectious agent to be able to vary by stock which is analogous to a random slopes and intercept structure like "+ (infectious agent |stock)". We also probably want to account for the non-independence of observations within years given that the current dataset applies a single common measure of infection to all stocks in a given year; this is analogous to a random intercept structure like "+ (1|year)".

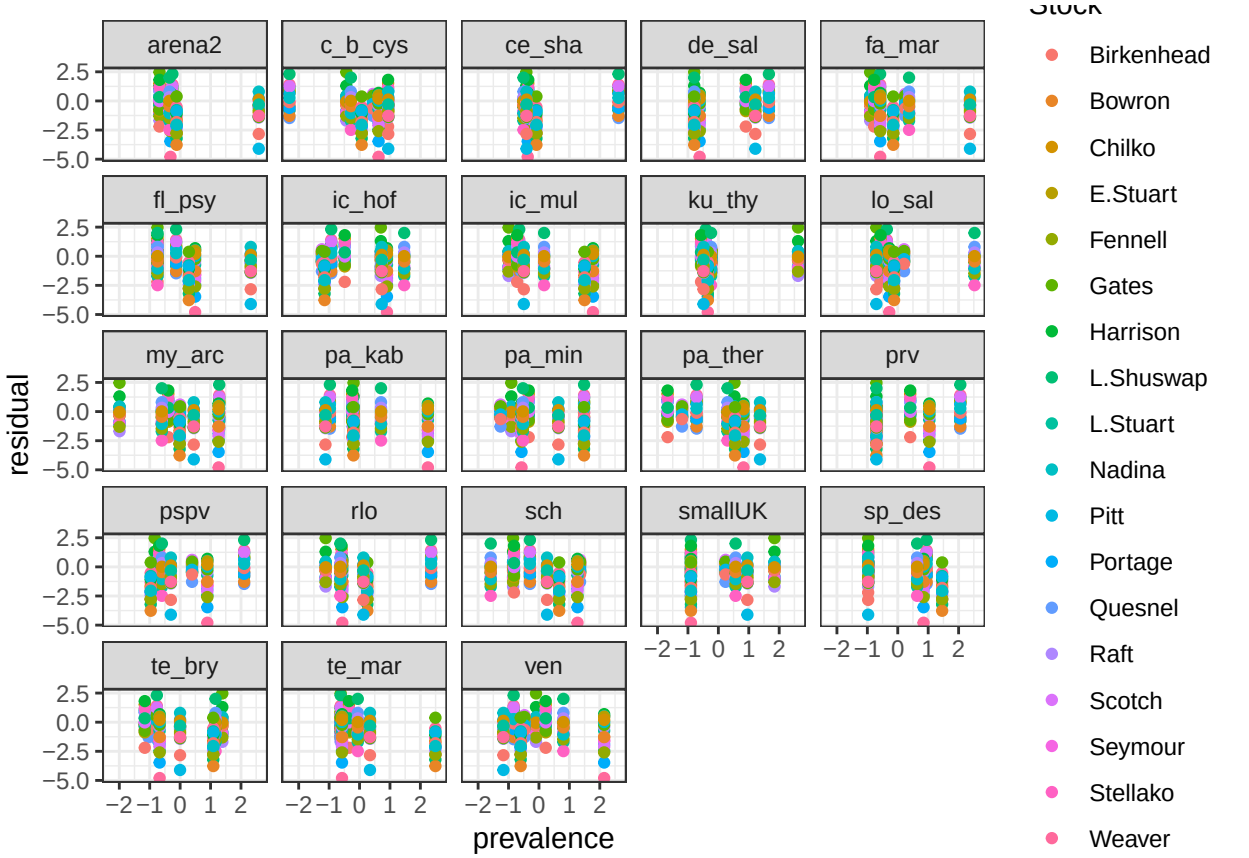
3. Lastly we have to decide whether or not to fit a single model to all agents simultaneously, or do this iteratively by stock. Initial model runs suggested that the single global model had a hard time converging. So for now let's stick to independent models for each agent.

Prevalence

w/ lme4

plot raw data

```
ggplot(inf_agt_resid_data,aes(prev_std, resid_value, color=Stock))+
    geom_point()+
    facet_wrap(~ agent,nrow=5)+
  xlab("prevalence")+
  ylab("residual")+
  theme_bw()
```



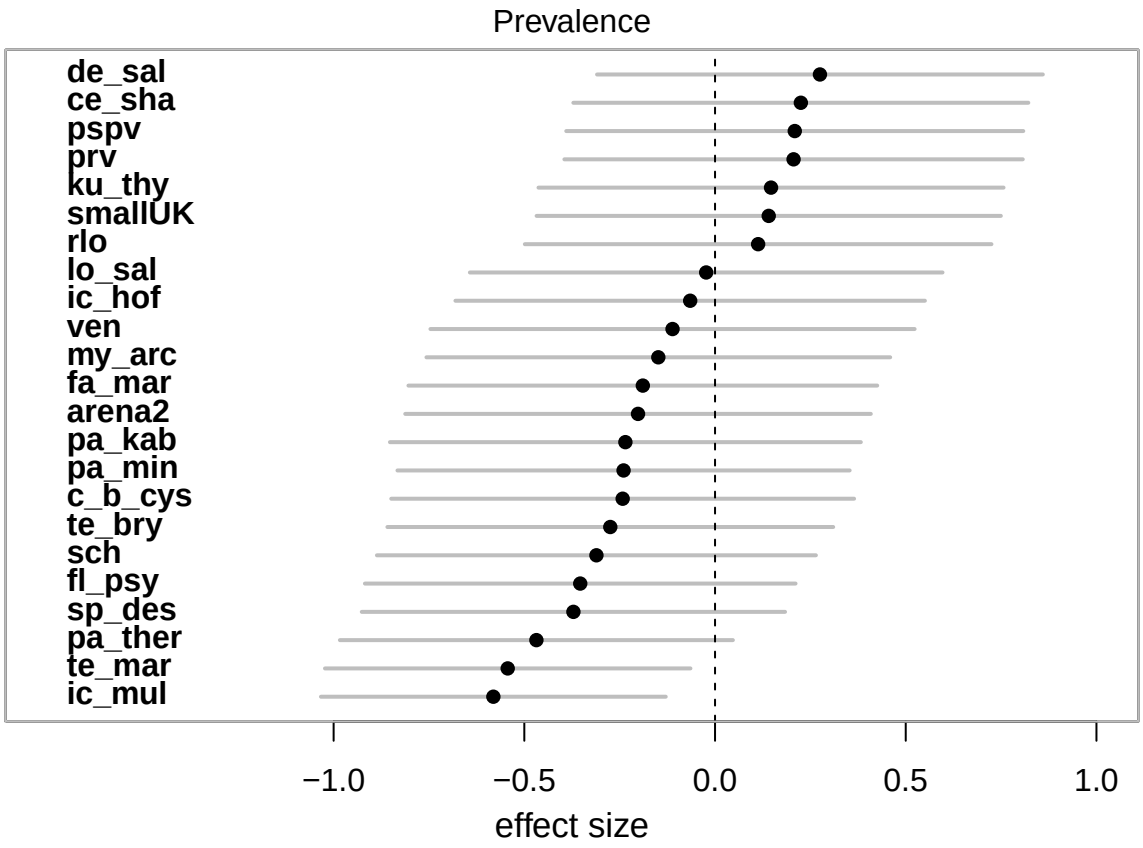
loop through agents fitting individual models

```
coefs <-matrix(NA,
               nrow = length(agents),
               ncol = 2,
               dimnames = list(agents,c("est","se")))

for(i in agents){
    data <- subset(inf_agt_resid_data, agent==i)
    model_ind <- lmer(resid_value ~ 0 + prev_std + (prev_std |Stock_Analysis)+(1|Year),
                      data = data,
                      REML = F)
    coefs[i,] <- summary(model_ind)$coefficients[1, 1:2]
  }

coefs_order <- coefs[order(coefs[,1]),]
```

plot effect sizes

error bars are +/- 2 x SE

```
par(mfrow=c(1,1), mar=c(3,1,1,1),oma=c(0.5,0.5,0.5,0.5))

plotCI(x = coefs_order[,1],
    y = seq(1,length(agents)),
    ui = (coefs_order[,1]+(coefs_order[,2]*2)),
    li = (coefs_order[,1]-(coefs_order[,2]*2)),
    err = "x",
    sfrac = 0 ,
    gap = 0,
    yaxt = "n",
    xaxt = "n",
    ylab = "",
    xlab = "",
    xlim = c(-1.75,1),
    pch = 16,
    lwd = 2,
    scol = "grey")

text(rep(-1.75,length(agents)),
    seq(1,length(agents)),
    labels = rownames(coefs_order),
    pos =4,
    font = 2)

axis(1, at = c(-1, -0.5, 0, 0.5, 1))
abline(v = 0, lty = 2)
box(col="grey")

mtext("effect size",1,line=2.2, cex=1.1)
mtext("Prevalence",3,line=0.25)
```



w/ STAN

loop through agents fitting individual models

change "eval = FALSE" in Rmd doc to actually run this code; takes 10-20 min.

```
coefs_stan <- matrix(NA,
                     nrow = length(agents),
                     ncol = 5,
                     dimnames = list(agents,c("lower","25","mid","75","upper")))

agents <- unique(inf_agt_resid_data$agent)

for(i in agents){
    data <- subset(inf_agt_resid_data, agent==i)
    model_ind_stan <- stan_lmer(resid_value ~ 0 + prev_std + (prev_std|Stock_Analysis) +(1|Year),
                                data = data,
                                REML = F)
    ind_coef <- summary(model_ind_stan,
                        pars = c("prev_std"),
                        probs = c(0.025,0.25,0.5,0.75, 0.975),
                        digits = 2)
    coefs_stan[i,] <- ind_coef[1,c(4:8)]
  }

write.csv(coefs_stan, "prev_coefs_stan.csv")
```

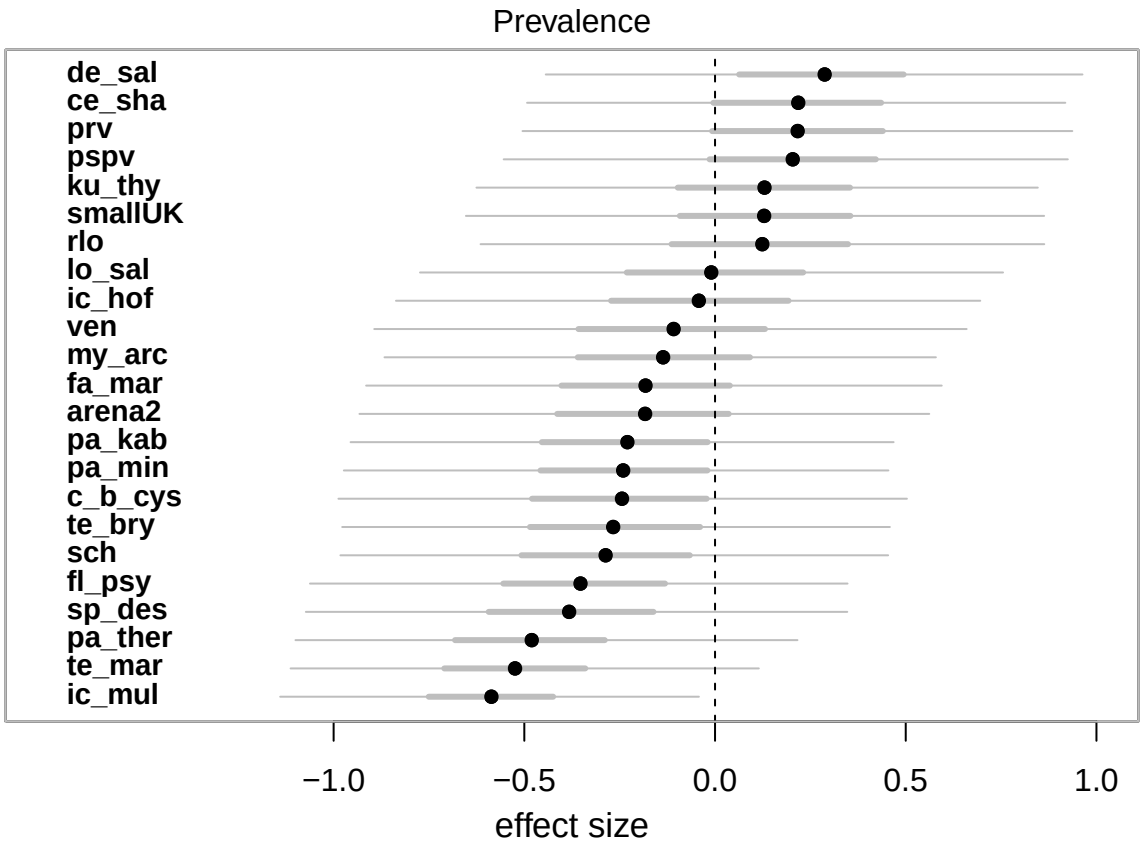
load parameter estimates if models above were not run.

```
coefs_stan <- read.csv("prev_coefs_stan.csv")
rownames(coefs_stan) <- coefs_stan[,1]
coefs_stan <- coefs_stan[,-1] # drop first column with agent names
```

plot effect sizes

error bars are 50% (thick grey) and 95% (thin grey) credibe interval

```
coefs_order <- coefs_stan[order(coefs_stan[,3]),]

par(mfrow=c(1,1), mar=c(3,1,1,1),oma=c(0.5,0.5,0.5,0.5))

plotCI(x = coefs_order[,3],
        y = seq(1,length(agents)),
        li = (coefs_order[,1]),
        ui = (coefs_order[,5]),
        err = "x",
        sfrac = 0 ,
        gap = 0,
        yaxt = "n",
        xaxt = "n",
        ylab = "",
        xlab = "",
        xlim = c(-1.75,1),
        pch = 16,
        scol = "grey")

plotCI(x = coefs_order[,3],
        y = seq(1,length(agents)),
        li = (coefs_order[,2]),
        ui = (coefs_order[,4]),
        err = "x",
        sfrac = 0 ,
        gap = 0,
        pch = 16,
        add = TRUE,
        lwd = 3,
        scol = "grey")

text(rep(-1.75,length(agents)),
        seq(1,length(agents)),
        labels = rownames(coefs_order),
        pos = 4,
        font = 2,
        cex=0.95)

axis(1, at = c(-1, -0.5, 0, 0.5, 1))
abline(v = 0, lty = 2)
box(col="grey")
mtext("effect size",1,line=2.2, cex=1.1)
mtext("Prevalence",3,line=0.25)
```

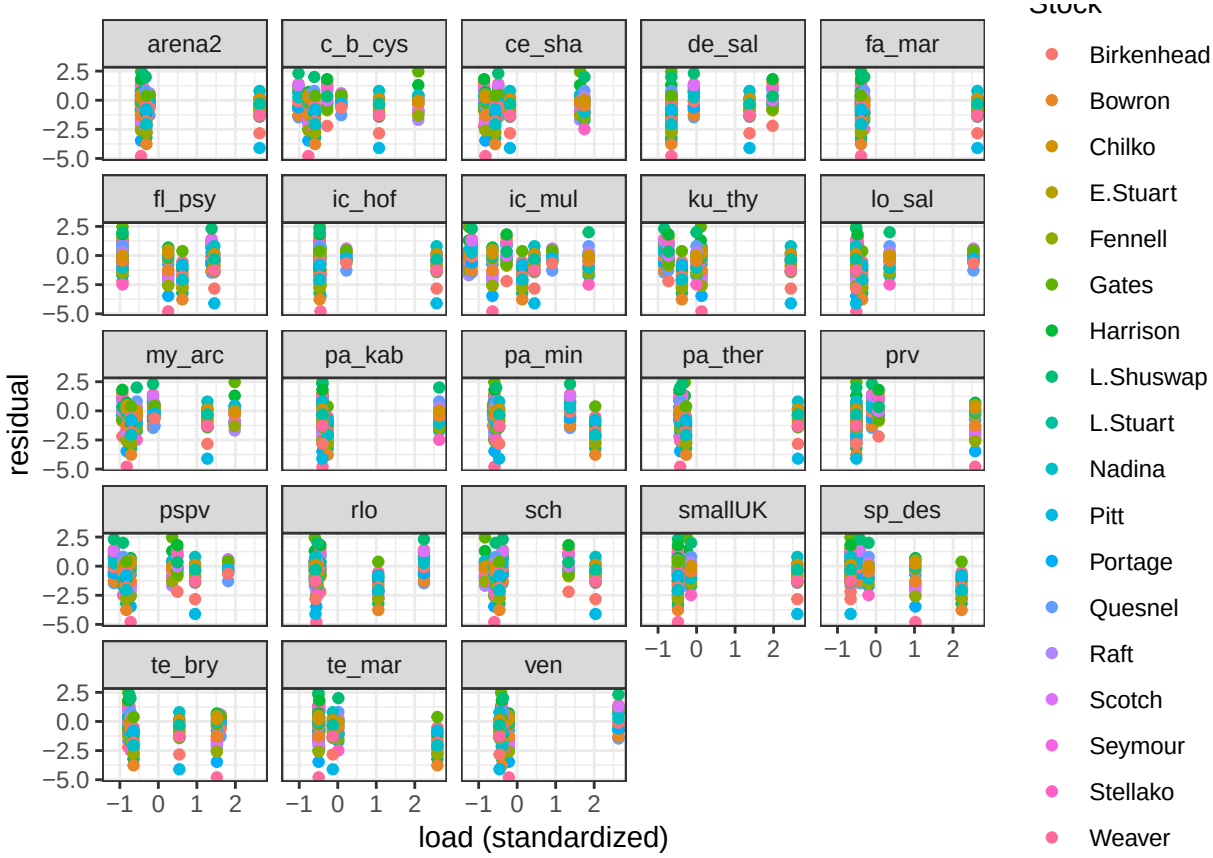


Intensity

w/ lme4

plot raw data

```
ggplot(inf_agt_resid_data,aes(load_std, resid_value, color=Stock))+
    geom_point()+
    facet_wrap(~ agent,nrow=5)+
    xlab("load (standardized)")+
    ylab("residual")+
    theme_bw()
```



loop through agents fitting individual models

```
coefs <-matrix(NA,
               nrow = length(agents),
               ncol = 2,
               dimnames = list(agents,c("est","se")))

agents <- unique(inf_agt_resid_data$agent)

for(i in agents){
    data <- subset(inf_agt_resid_data, agent==i)
    model_ind <- lmer(resid_value ~ 0 + load_std + (load_std |Stock_Analysis)+(1|Year),
                      data = data)
    coefs[i,] <- summary(model_ind)$coefficients[1, 1:2]
  }
```

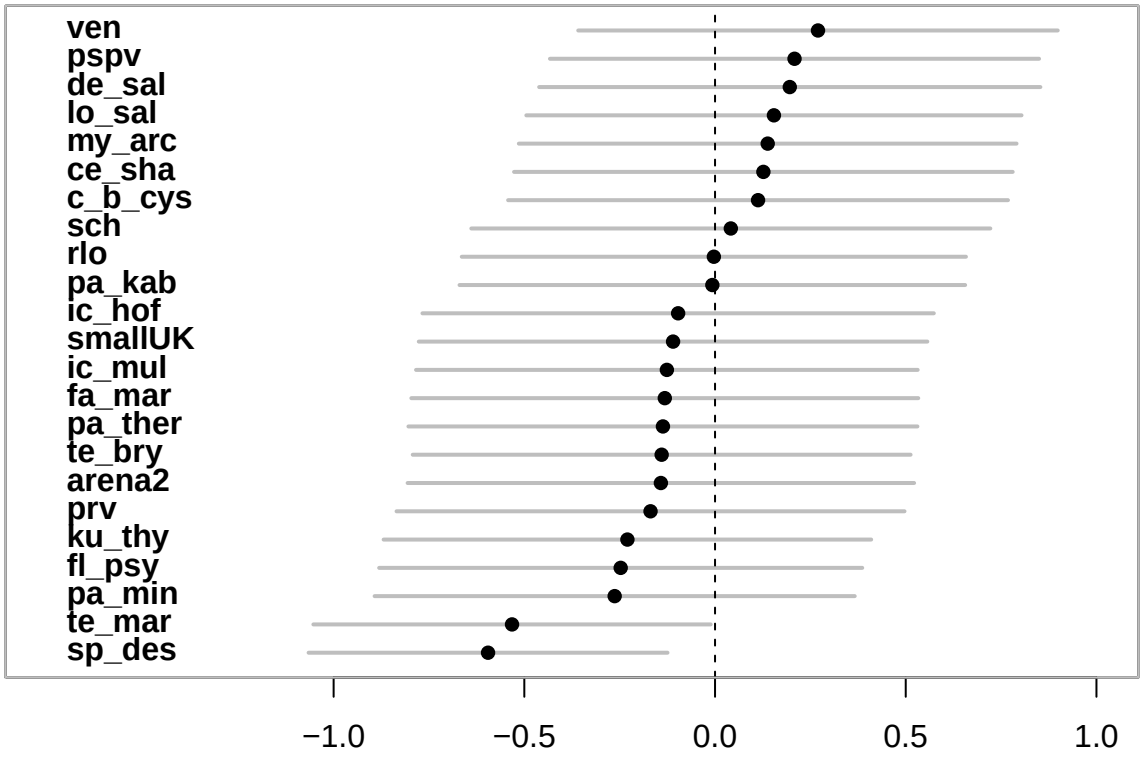
```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =
## control$checkConv, : Model failed to converge: degenerate Hessian with 1
## negative eigenvalues
```

```
coefs_order <- coefs[order(coefs[,1]),]
```

plot effect sizes

```r
par(mfrow=c(1,1), mar=c(3,1,1,1),oma=c(0.5,0.5,0.5,0.5))

plotCI(x = coefs_order[,1],
        y = seq(1,length(agents)),
        ui = (coefs_order[,1]+(coefs_order[,2]*2)),
        li = (coefs_order[,1]-(coefs_order[,2]*2)),
        err = "x",
        sfrac = 0 ,
        gap = 0,
        yaxt = "n",
        xaxt = "n",
        ylab = "",
        xlab = "effect size",
        xlim = c(-1.75,1),
        pch = 16,
        lwd= 2,
        scol = "grey")

text(rep(-1.75,length(agents)),
    seq(1,length(agents)),
    labels = rownames(coefs_order),
    pos = 4,font = 2)

axis(1, at = c(-1, -0.5, 0, 0.5, 1))
abline(v = 0, lty = 2)
box(col="grey")
```



w/ STAN

loop through agents fitting individual models

change "eval = FALSE" in Rmd doc to actually run this code; takes 10-20 min.

```
coefs_stan <-matrix(NA,
                nrow = length(agents),
                ncol = 5,
                dimnames = list(agents,c("lower","25","mid","75","upper")))

agents <- unique(inf_agt_resid_data$agent)

for(i in agents){
    data <- subset(inf_agt_resid_data, agent==i)
    model_ind_stan <- stan_lmer(resid_value ~ 0 + load_std + (load_std |Stock_Analysis) + (1|Year),
                                data = data)

    ind_coef <- summary(model_ind_stan,
                        pars = c("load_std"),
                        probs = c(0.025,0.25,0.5,0.75, 0.975),
                        digits = 2)

    coefs_stan[i,] <- ind_coef[1,c(4:8)]
  }
```

```
write.csv(coefs_stan, "int_coefs_stan.csv")
```

load parameter estimates if models above were not run.

```
coefs_stan <- read.csv("int_coefs_stan.csv")
rownames(coefs_stan) <- coefs_stan[,1]
coefs_stan <- coefs_stan[,-1] # drop first column with agent names
```

plot effect sizes

error bars are 50% (thick grey) and 95% (thin grey) credibe interval

```
coefs_order <- coefs_stan[order(coefs_stan[,3]),]

par(mfrow=c(1,1), mar=c(3,1,1,1),oma=c(0.5,0.5,0.5,0.5))

plotCI(x = coefs_order[,3],
        y = seq(1,length(agents)),
        li = (coefs_order[,1]),
        ui = (coefs_order[,5]),
        err = "x",
        sfrac = 0 ,
        gap = 0,
        yaxt = "n",
        xaxt = "n",
        ylab = "",
        xlab = "",
        xlim = c(-1.75,1),
        pch = 16,
        scol = "grey")

plotCI(x = coefs_order[,3],
        y = seq(1,length(agents)),
        li = (coefs_order[,2]),
        ui = (coefs_order[,4]),
        err = "x",
        sfrac = 0 ,
        gap = 0,
        pch = 16,
        add = TRUE,
        lwd = 3,
        scol = "grey")

text(rep(-1.75,length(agents)),
        seq(1,length(agents)),
        labels = rownames(coefs_order),
        pos = 4,
        font = 2,
        cex=0.95)

axis(1, at = c(-1, -0.5, 0, 0.5, 1))
abline(v = 0, lty = 2)
box(col="grey")
mtext("effect size",1,line=2.2, cex=1.1)
```

```
mtext("Intensity",3,line=0.25)
```