

Final Paper Draft

Karan

April 17, 2018

The study (Persha, Agrawal, and Chhatre 2011) that I am reproducing examines biodiversity conservation and forest-based livelihood outcomes and their relationship with specific contextual factors; specifically: Forest Size, Community participation in rule making, and commercial dependence on forests. The study uses a data set of 84 forest sites from six countries in East Africa and South Asia. This data set was compiled as part of the International Forestry Resources and Institutions (IFRI) research programme. The data set captures social, ecological and governance data from forests in human-dominated landscapes, of the 84 cases, 30 are drawn from East Africa (Kenya:6, Tanzania:7, and Uganda:3) and 54 are from South Asia (Bhutan:2, India:27, Nepal:25).

Before assessing the methods applied in this paper, we shall first discuss the significance of the IFRI database within research on collective action and the management of natural resources. The IFRI program grew out of a concern to address the many knowledge gaps that exist in our understanding of how human choices impact processes of forest change (Ostrom 1998). This field of research has been largely characterized by a case study based approach and while large N-studies are an appropriate approach for providing generalized findings in related to the management of common pool resources, they have been very hard to conduct (Poteete, Janssen, and Ostrom 2010). Attempts at doing large N work relied on Meta-analysis, but can be confounded by inconsistent coding, missing data, and sampling issues.

IFRI aims to collect comparable data in numerous sites that will assist communities, policy makers, and scientists in analyzing and addressing the factors that shape human action and forest conditions, and to provide policy recommendations based on comparative analyses (Ostrom and Wertime 1998). It lends itself to testing the numerous models and hypotheses that exist concerning the relationships between humans and forests. By including a wide range of questions drawn from many disciplinary perspectives, IFRI has the potential to address new questions as they emerge. This program is among the first to collect reliable forest measurements from plot data in association with systematic, detailed data collection on socioeconomic, demographic, institutional, and biophysical characteristics in multiple sites across time.

The full paper is available here: <https://doi.org/10.1126/science.1199343>

The data set is available here: <http://www.ifriresearch.net/resources/data/>

Dependent Variables

The authors construct the dependent/outcome variable from measures used as indicators of tree species richness and dependence on the forest for subsistence livelihoods. Using standardized z-scores of these variables they classify each case into one of three joint outcome categories based on whether the standardized value for each case was greater than or less than zero (i.e. the mean) within each forest type. Thus they create three outcome categories High-High, Low-low, and Trade-off (See table 1)

	Outcome	Tree Species	Subsistence Livelihood	Description
1	High-High	>0	>0	Sustainable forest Systems
2	Low-Low	<0	<0	Unsustainable forest Systems
3	Trade-off	>0	<0	Trade-off between Livelihoods and Forest sustainability
4	Trade-off	<0	>0	Trade-off between Livelihoods and Forest sustainability

Table 1: Levels of Joint Outcome Variable

In the next section I summarize the constituent parts of the outcome variables from the data set. As mentioned above the first outcome of interest in this study is forest biodiversity. The authors use a non-parametric

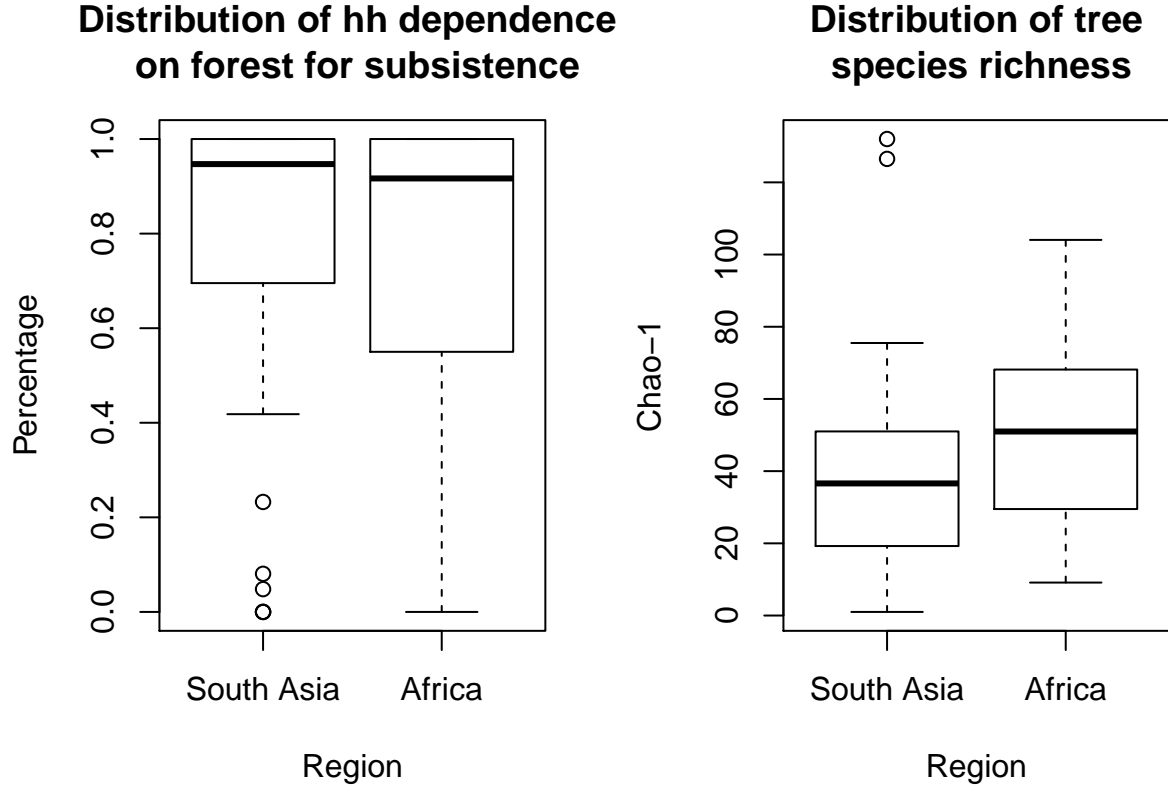
	South Asia	Africa
Low-Low	9	2
Tradeoff	29	21
High-High	16	7

Table 2: Distribution of Joint Outcome Variable

Chao1 estimator of species richness (Chao et al. 2005) as a proxy indicator for overall forest biodiversity value. The second outcome variable of interest is an estimate of the percent of households who depend significantly on the forest for subsistence household needs. I create box plots that summarize the outcome variables for each region.

	Median	Mean	Min.	Max.	sigma	Coverage(0)	Coverage(97.5)
Tree Species (Chao-1)	38.82	43.38	1.00	132.00	26.50	1.00	103.32
Subsistence Livelihoods	0.95	0.78	0.00	1.00	26.50	0.00	1.00

Table 3: Summary of Outcome Variables



Independent Variables

The paper uses 3 independent/explanatory variables identified from literature on biodiversity and livelihoods

Forest size (fsize): Refers to the administrative area under a given management designation rather than the ecological extent of a forest. Forest rule-making by local forest users(gmakerule_rec): This is a binary variable that records whether local forest user participate in forest rule-making (1=yes, 0=no) dependence

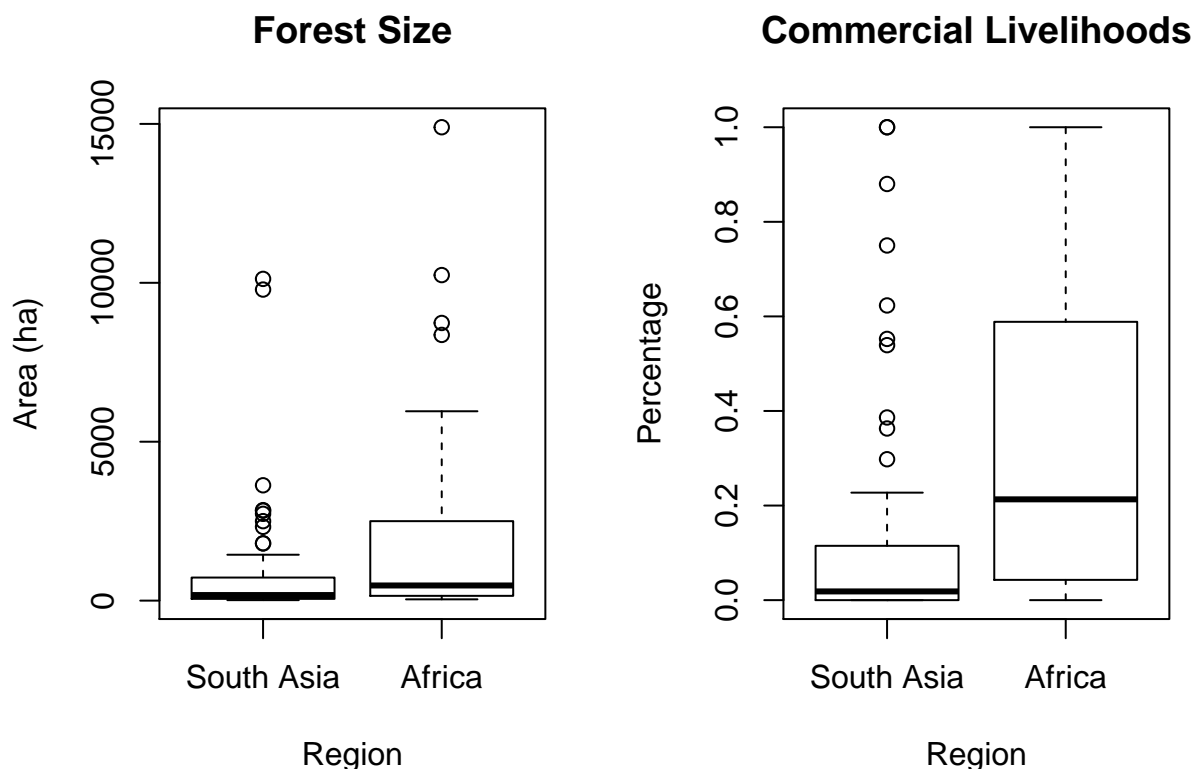
on forest for commercial livelihoods: An estimate of the percent of households who depend on the forest for household cash income derived through commercial forest activities.

	Median	Mean	Min.	Max.	sigma	Coverage(2.5)	Coverage(97.5)
Forest Size	271.19	1421.90	15.00	14895.00	2783.76	15.00	10097.87
Commercial Livelihoods	0.05	0.21	0.00	1.00	0.30	0.00	1.00

Table 4: Summary of Independent Variables

	South Asia	Africa
0	32	21
1	22	9

Table 5: Participation in Rulemaking across Region



Model

The authors use ordered logistic regression analysis to estimate the effects of three independent variables on the probability of obtaining each of the three joint outcome categories they define. They do not specify the reason for choosing an ordered logistic regression over others. They use their model with bootstrapped standard errors (1000 replications). The authors focus on marginal effects rather than the regression coefficients. I use the `oglmx` package to run the ordered logistic regression. Marginal effects are calculated holding rule making participation at its median $IE = 0$ and all other variables at their mean.

Justification for Using Ordinal Logit

What could the justification for the model selected be. This section would assess this question, going beyond the claim that the we have and ordered outcome variable. However given that the authors specifically constructed an ordered outcome variable, they must have created this with in mind.

Why do the authors use Marginal effects over the regression coefficients?? - Provide justification

Bootstrapped Confidence Intervals

In this section I shall reproduce the standard errors that the authors produce. The autos do not specify whether they use a parametric or non-parametric bootstrap. It would have been helpful if they had sited a source for their bootstrap protocol. Calculating ordered logistic regression coefficients, marginal effects, and bootstrapped confidence intervals in stata is much more streamlined than it is in R. Confidence intervals were generated here using a bootstrap over 10000 simulations(with replacement).

	Marg. Eff	Std. error	t value	Pr(> t)	2.5%	97.5%
Rulemaking Participation	-0.118735	0.044975	-2.639998	0.008291	-0.235409	-0.010699
Forest Size	-0.024386	0.013085	-1.863599	0.062378	-0.052547	-0.001814
Commercial Livelihoods	-0.147156	0.078823	-1.866908	0.061914	-0.252665	-0.034929

Table 6: Joint Outcome Category (Low-Low): Marginal effects of ordered Logit Regression

	Marg. Eff	Std. error	t value	Pr(> t)	2.5%	97.5%
Rulemaking Participation	-0.157842	0.082364	-1.916412	0.055313	-0.271367	0.032879
Forest Size	-0.023048	0.015419	-1.494715	0.134989	-0.039749	0.005579
Commercial Livelihoods	-0.139083	0.092023	-1.511397	0.130687	-0.370600	-0.009903

Table 7: Joint Outcome Category (Trade-off): Marginal effects of ordered Logit Regression

	Marg. Eff	Std. error	t value	Pr(> t)	2.5%	97.5%
Rulemaking Participation	0.276577	0.103957	2.660492	0.007803	0.032006	0.487997
Forest Size	0.047433	0.023992	1.977012	0.048040	0.006630	0.081644
Commercial Livelihoods	0.286239	0.143276	1.997809	0.045737	0.068734	0.392828

Table 8: Joint Outcome Category (High-High): Marginal effects of ordered Logit Regression

Model Assessment

The authors conduct a series of post estimation diagnostics to check for violations of model assumptions. They make the following claims in their analysis. Here I reproduce their analysis of what they define as least well predicted cases and the effect that these have on their analysis. The problem as they identify it is that for 11 of the cases the indicators of tree species richness and livelihoods are very close (above or below) to the average. This raises the problem that they may have misclassified the outcome category in these cases. To address this issue they reclassify these cases by changing the sign on these cases and then reclassifying the outcome categories. They then re-ran the ordered logit regression on the full data set, using the reclassified values in the dependent variable for these eleven cases, and assessed differences in the marginal effects and predicted probabilities for each independent variable as obtained from the reclassified dependent variable.

We consider the cases where the zscores we calcuted are between $-1 < \text{treespecies richness/subsistence livelihoods} < .1$. For these cases we switch the signs and then recategorise the cases. As can be seen here the signs of the marginal effects that were calculated do not change when least well predicted cases are accounted for.

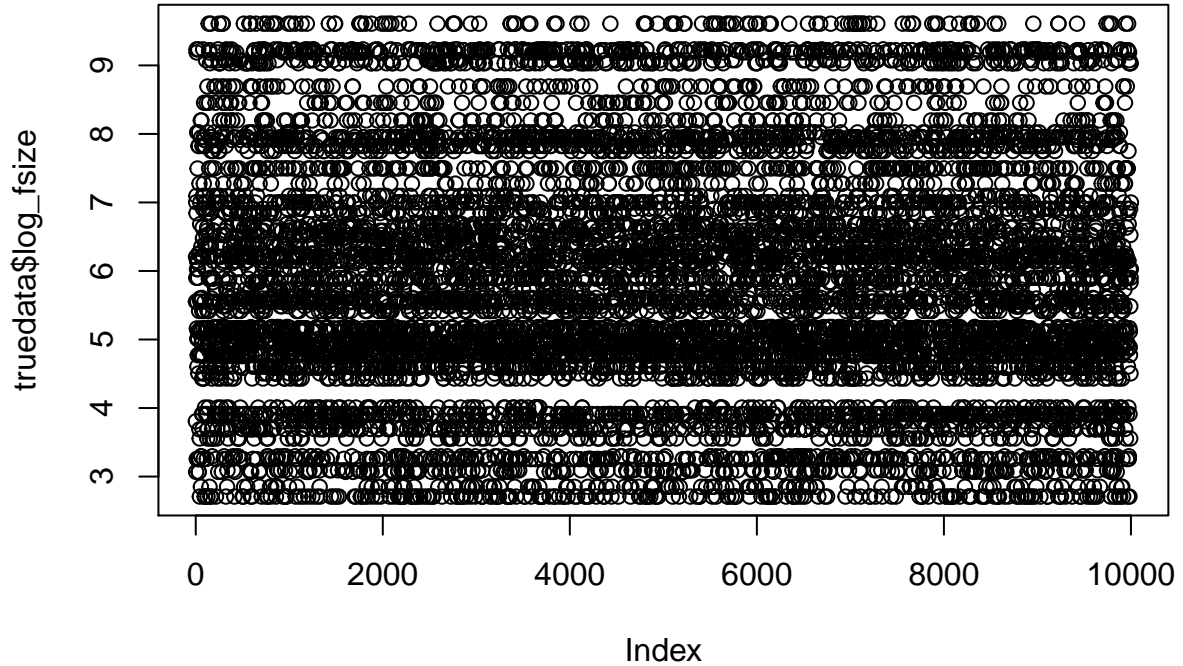
	Rulemaking Participation	Forest Size	Commercial Livelihoods
Low-Low	-0.119	-0.024	-0.147
Trade-off	-0.158	-0.023	-0.139
High-High	0.277	0.047	0.286

Table 9: Marginal Effects from observed data

	Rulemaking Participation	Forest Size	Commercial Livelihoods
Low-Low	-0.112	-0.018	-0.230
Trade-off	-0.075	-0.008	-0.097
High-High	0.187	0.025	0.327

Table 10: Marginal Effects testing for poorly defined outcomes

In this section I test the reliability of Multinomial Logistic regression in small samples using simulations. I will use the method outlined in Epstein et al (Epstein et al. 2014) to determine the parameter sign change, Type 1 and Type 1 error rates. Briefly the method I will follow is: First, simulate a hypothetical population of 10,000 observations, while maintaining the relationship between the dependent and independent variables. This method changes the distribution of the dependent & independent variables. However this is not a problem as we are doing this to check how well ordered logistic regression performs under when n is relatively small ie <100.



I use an Ordered logistic regression on the full sample of 10,000 cases to record the “true” sign and statistical significance of the marginal effects. We then take 100 separate draws on 84 cases each (without replacement) and calculate the marginal effects. While this is similar to the bootstrap, it differs in that we sample without replacement, this means that each observation can only appear once in our sample. We assess the no. of

times the sign of the marginal effect changes, the maximum & minimum and means for each of the pseudo random samples. from the table below we can see that sign changes is relatively infrequent.

In order to address the likelihood of type 1 errors we first take our true data set and shuffle the dependent variable. Now there is by definition no relationship between the dependent and independent variables. We calculate the Marginal effects for this data set and then conduct draw 100 pseudo random draws of 80 observations, using the standard errors and p values generated by the package (using delta method) we calculate the type one error rate, ie the likelihood of rejecting the null hypothesis when it is true. From the table we see that the false positive rate is generally under 5% at $\alpha = .05$. In addition we conduct a

	Change Sign	in	Simulation: Type 1 Error Rate (0.05)	Simulation: Type 1 Error Rate (0.1)	Permutation: Type 1 Error Rate (0.05)	Permutation: Type 1 Error Rate (0.5)
Rulemaking Participation	0.005		0.032	0.082	0.042	0.083
Forest Size	0.011		0.040	0.104	0.042	0.079
Commercial Livelihoods	0.020		0.037	0.099	0.055	0.108

Table 11: Joint Outcome Category (Low-Low): Marginal effects of ordered Logit Regression

% latex table generated in R 3.5.0 by xtable 1.8-2 package % Wed May 16 15:36:00 2018

	Change Sign	in	Simulation: Type 1 Error Rate (0.05)	Simulation: Type 1 Error Rate (0.1)	Permutation: Type 1 Error Rate (0.05)	Permutation: Type 1 Error Rate (0.5)
Rulemaking Participation	0.009		0.014	0.043	0.002	0.012
Forest Size	0.030		0.011	0.032	0.000	0.012
Commercial Livelihoods	0.039		0.007	0.034	0.000	0.018

Table 12: Joint Outcome Category (Trade-off): Marginal effects of ordered Logit Regression

	Change Sign	in	Simulation: Type 1 Error Rate (0.05)	Simulation: Type 1 Error Rate (0.1)	Permutation: Type 1 Error Rate (0.05)	Permutation: Type 1 Error Rate (0.5)
Rulemaking Participation	0.009		0.014	0.043	0.002	0.012
Forest Size	0.030		0.011	0.032	0.000	0.012
Commercial Livelihoods	0.039		0.007	0.034	0.000	0.018

Table 13: Joint Outcome Category (Trade-off): Marginal effects of ordered Logit Regression

References

```
knitr::opts_chunk$set(echo = TRUE)

##install.packages("permute")
##install.packages("oglmx")
##install.packages("dplyr")
##install.packages("xtable")
##install.packages("haven")
##install.packages("knitr")
##install.packages("doParallel")
library(permute)
library(oglmx)
library(dplyr)
library(xtable)
library(haven)
```

```

library(knitr)
library(foreach)
library(doParallel)

##Import the persha et al 2011 data set available here: http://www.ifriresearch.net/resources/data/

##Paper available here: https://doi.org/10.1126/science.1199343

fpdat <- read_dta(
  "C:/Users/karan/Desktop/PS 531/pershaagrawalchhatre_jan_2011 (2)/PershaAgrawalChhatre_Science2011.dta"
)
View(fpdat)
o <- c("High-High", "Low-Low", "Trade-off", "Trade-off")
t <- c(">0", "<0", ">0", "<0")
s <- c(">0", "<0", "<0", ">0")
d <- c("Sustainable forest Systems", "Unsustainable forest Systems", "Trade-off between Livelihoods and

tabledesc <- as.data.frame(cbind(o,t,s,d))
colnames(tabledesc) <- c("Outcome", "Tree Species", "Subsistence Livelihood", "Description")

table1 <- print(xtable(tabledesc, caption="Levels of Joint Outcome Variable"),
  comment = FALSE)
## Assign value labels to the outcome variable of interest

fpdat$Region1 <- factor(fpdat$Region,
  levels = c(0,1),
  labels = c("South Asia", "Africa"))

fpdat$cat_zsubchao_3cat1 <- factor(fpdat$cat_zsubchao_3cat,
  levels = c(1,2,3),
  labels = c("Low-Low", "Tradeoff", "High-High"))

## Frequency table
tableoutfreq <- table(fpdat$cat_zsubchao_3cat1, fpdat$Region1 )

table2 <- print(xtable(tableoutfreq, caption="Distribution of Joint Outcome Variable"),
  comment = FALSE)

##As the outcome variable is composed of two other variables we
##summarise the standardised and non standardised version of
##these variables ie the measure of subsistence livelihoods and tree species richness

outcomesummary <- rbind(c(summary(fpdat$chao_1_mean)[c(1,3,4,6)], sd(fpdat$chao_1_mean),
  quantile(fpdat$chao_1_mean, probs=c(.0, .975))),
  c(summary(fpdat$avgofghhsubsis_pcthh)[c(1,3,4,6)], sd(fpdat$chao_1_mean),
  quantile(fpdat$avgofghhsubsis_pcthh, probs=c(.0, .975))))

colnames(outcomesummary)[c(5,6,7)] <- c("sigma", "Coverage(0)", "Coverage(97.5)")

```

```

rownames(outcomesummary) <- c("Tree Species (Chao-1)", "Subsistence Livelihoods")

outcomesummary <- outcomesummary[,c(2,3,1,4,5,6,7)]

table3 <- print(xtable(outcomesummary, caption = "Summary of Outcome Variables"),
  comment = FALSE)

par(mfrow = c(1, 2))

boxplot(avgofghhsusis_pcthh~Region1, data=fpdat,
  main= "Distribution of hh dependence \n on forest for subsistence",
  xlab="Region", ylab="Percentage")
boxplot(chao_1_mean ~ Region1, data=fpdat,
  main= "Distribution of tree\n species richness",
  xlab="Region", ylab="Chao-1")

indsummary <-rbind(c(summary(fpdat$fsize)[c(1,3,4,6)],sd(fpdat$fsize),
  quantile(fpdat$fsize,probs=c(.0,.975))),
  c(summary(fpdat$avgofghhcomm_pcthh)[c(1,3,4,6)], sd(fpdat$avgofghhcomm_pcthh),
  quantile(fpdat$avgofghhcomm_pcthh,probs=c(.025,.975))))

colnames(indsummary)[c(5,6,7)] <- c("sigma", "Coverage(2.5)", "Coverage(97.5)")

rownames(indsummary) <- c("Forest Size", "Commercial Livelihoods")

indsummary <- indsummary[,c(2,3,1,4,5,6,7)]

table4 <- print(xtable(indsummary,caption="Summary of Independent Variables"),
  comment = FALSE)

## Rule-making
tableinfreq <- table(fpdat$gmakerule_rec, fpdat$Region1)

table5 <- print(xtable(tableinfreq, caption="Participation in Rulemaking across Region"),
  comment = FALSE)

par(mfrow = c(1, 2))
boxplot(fsize ~ Region1, data=fpdat, main= "Forest Size", xlab="Region", ylab="Area (ha)")
boxplot(avgofghhcomm_pcthh ~ Region1, data=fpdat,
  main= "Commercial Livelihoods", xlab="Region", ylab="Percentage")

##install.packages("ogmlx")

repdat <- fpdat[,c("cat_zsubchao_3cat","gmakerule_rec", "log_fsize", "avgofghhcomm_pcthh")]

as.ordered(repdat$cat_zsubchao_3cat)
as.factor(repdat$gmakerule_rec)

#vglmFit <- vglm(cat_zsubchao_3cat ~ gmakerule_rec + log_fsize + avgofghhcomm_pcthh, family=propodds, d

```



```

model <- as.ordered(cat_zsubchao_3cat) ~ log_fsize + as.factor(gmakerule_rec) + avgofghhcomm_pcthh

##reg <- ologit.reg(model, newdata, start = NULL, weights=NULL, beta = NULL, threshparam = NULL,
  ##analhessian = TRUE, na.exclude, savemodelframe = FALSE, robust = FALSE, Force = FALSE)

reg<-oglmx(model, data=repdat, link="logit", constantMEAN = FALSE,
  constantSD = FALSE,delta=0,threshparam = NULL)

#dummyzero = TRUE, takes the value of binary value to be zero this is consistent with

me <- margins.oglmx(reg, Vars = NULL, outcomes = "All", atmeans = TRUE, AME = FALSE,
  ascontinuous = FALSE, location = NULL)

bl <- rbind((me[[1]])[,1], (me[[2]])[,1], (me[[3]])[,1])

##Lets do our bootstrap. First we create a function & then we replicate

mybs<- function(datx){
bsdata<- sample_n(datx, size=84,replace=TRUE)
bsreg <- oglmx(model, data=bsdata, link="logit", constantMEAN = FALSE,
  constantSD = FALSE,delta=0,threshparam = NULL)
bsme <- margins.oglmx(bsreg, Vars = NULL, outcomes = "All", atmeans = TRUE, AME = FALSE,
  ascontinuous = FALSE, location = NULL)
sumbs <- t(c((bsme[[1]])[,1], (bsme[[2]])[,1], (bsme[[3]])[,1]))

##test <- replicate(nsamps, mybs(newdata))
return(sumbs)
}

#library(doParallel)
#cl <- makeCluster(detectCores(), type='PSOCK')
#registerDoParallel(cl)

set.seed(143435)
test <- replicate(10, mybs(repdat))

#Calculate the bootstrap confidence intervals
cifsizell <-quantile(test[,2,],c(.025,.975)) ## percentile bootstrap interval
cifsizelh <- quantile(test[,8,],c(.025,.975)) ## percentile bootstrap interval
cifsizeto <- quantile(test[,5,],c(.025,.975)) ## percentile bootstrap interval

cirulell <-quantile(test[,1,],c(.025,.975)) ## percentile bootstrap interval
cirulehh<-quantile(test[,7,],c(.025,.975)) ## percentile bootstrap interval
ciruleto <-quantile(test[,4,],c(.025,.975)) ## percentile bootstrap interval

circomll <-quantile(test[,3,],c(.025,.975)) ## percentile bootstrap interval
circomhh<-quantile(test[,9,],c(.025,.975)) ## percentile bootstrap interval
circomto <-quantile(test[,6,],c(.025,.975)) ## percentile bootstrap interval

```

```

##reg

confi <- rbind(cifsize11, cirule11, circom11, cifsize10, circom10, cirule10, cifsizehh, cirulehh, circomhh)

rownames(me[[1]]) <- c("Rulemaking Participation", "Forest Size", "Commercial Livelihoods")
rownames(me[[2]]) <- c("Rulemaking Participation", "Forest Size", "Commercial Livelihoods")
rownames(me[[3]]) <- c("Rulemaking Participation", "Forest Size", "Commercial Livelihoods")

me[[1]] <- cbind(me[[1]], confi[c(2,1,3),])
me[[2]] <- cbind(me[[2]], confi[c(5,4,6),])
me[[3]] <- cbind(me[[3]], confi[c(8,7,9),])

l1 <- as.data.frame(me[[1]])
to<- as.data.frame(me[[2]])
hh<- as.data.frame(me[[3]])

tablemell <- print(xtable(l1,caption="Joint Outcome Category (Low-Low): Marginal effects of ordered Logit",
                        digits=c(0,6,6,6,6,6,6)), comment = FALSE)

tablemeto <- print(xtable(to,caption="Joint Outcome Category (Trade-off): Marginal effects of ordered Logit",
                        digits=c(0,6,6,6,6,6,6)), comment = FALSE)

tablemehh <- print(xtable(hh,caption="Joint Outcome Category (High-High): Marginal effects of ordered Logit",
                        digits=c(0,6,6,6,6,6,6)), comment = FALSE)

casdat <- fpmat
casdat$z2HHsub_ft5all2 <- casdat$z2HHsub_ft5all
casdat$z2chao_ft5all2 <- casdat$z2chao_ft5all

casdat$z2HHsub_ft5all2 <- ifelse(casdat$z2HHsub_ft5all2 < -.1, casdat$z2HHsub_ft5all2,
                                ifelse(casdat$z2HHsub_ft5all2 > .1, casdat$z2HHsub_ft5all2,
                                        -1*casdat$z2HHsub_ft5all2))

casdat$z2chao_ft5all2 <- ifelse(casdat$z2chao_ft5all2 < -.1, casdat$z2chao_ft5all2,
                                ifelse(casdat$z2chao_ft5all2 > .1, casdat$z2chao_ft5all2,
                                        -1*casdat$z2chao_ft5all2))

casdat$newcat <- ifelse(casdat$z2HHsub_ft5all2 < 0 & casdat$z2chao_ft5all2 < 0, 1,
                        ifelse(casdat$z2HHsub_ft5all2 > 0 & casdat$z2chao_ft5all2 < 0, 2,
                                ifelse(casdat$z2HHsub_ft5all2 < 0 & casdat$z2chao_ft5all2 > 0, 2, 3)))

as.ordered(casdat$newcat)
as.factor(casdat$gmakerule_rec)

#vglmFit <- vglm(cat_zsubchao_3cat ~ gmakerule_rec + log_fsize + avgofghhcomm_pcthh, family=propodds, data=casdat)

modelcas <- as.ordered(newcat) ~ log_fsize + as.factor(gmakerule_rec) + avgofghhcomm_pcthh

```

```

##reg <- ologit.reg(model, newdata, start = NULL, weights=NULL, beta = NULL, threshparam = NULL,
                    ##analhessian = TRUE, na.exclude, savemodelframe = FALSE,
##robust = FALSE, Force = FALSE)

casreg<-oglmx(modelcas, data=casdat, link="logit", constantMEAN = FALSE,
              constantSD = FALSE,delta=0,threshparam = NULL)

#dummyzero = TRUE, takes the value of binary value to be zero this is consistent with

casme <- margins.oglmx(casreg, Vars = NULL, outcomes = "All", atmeans = TRUE, AME = FALSE,
                      ascontinuous = FALSE, location = NULL)

casbl <- rbind((casme[[1]])[,1], (casme[[2]])[,1], (casme[[3]])[,1])

rownames(casbl) <- c("Low-Low", "Trade-off", "High-High")

rownames(bl) <- c("Low-Low", "Trade-off", "High-High")

colnames(casbl) <- c("Rulemaking Participation", "Forest Size", "Commercial Livelihoods")

colnames(bl) <- c("Rulemaking Participation", "Forest Size", "Commercial Livelihoods")

tabbl <- print(xtable(bl, align = "cccc", latex.environments = "center",
                      caption="Marginal Effects from observed data",
                      digits=c(0,3,3,3)),comment = FALSE)

tabcasbl <- print(xtable(casbl, align = "cccc", latex.environments = "center",
                        caption="Marginal Effects testing for poorly defined outcomes",
                        digits=c(0,3,3,3)), comment = FALSE)

set.seed(454654)
truedata <- sample_n(repdat, size=10000, replace=TRUE)

truedata$log_fsizej<- jitter(truedata$log_fsize, factor = 1, amount = NULL)

model2 <- as.ordered(cat_zsubchao_3cat) ~ log_fsizej + as.factor(gmakerule_rec) + avgofghhcomm_pcthh

plot(truedata$log_fsize)

treg<-oglmx(model2, data=truedata, link="logit", constantMEAN = FALSE,
            constantSD = FALSE,delta=0,threshparam = NULL)

trueeme <- margins.oglmx(treg, Vars = NULL, outcomes = "All", atmeans = TRUE, AME = FALSE,
                        ascontinuous = FALSE, location = NULL)

##plot(density(truedata$log_fsize))
##plot(density(truedata$log_fsizej))

```

```

##plot(density(repdat$log_fsize))

##Lets do our bootstrap. First we create a function & then we replicate

mybs1<- function(datx){
bsdata<- sample_n(datx, size=84,replace= FALSE)
trreg <- oglmx(model2, data=bsdata, link="logit", constantMEAN = FALSE,
               constantSD = FALSE,delta=0,threshparam = NULL)
trme <- margins.oglmx(trreg, Vars = NULL, outcomes = "All", AME = FALSE,
                      ascontinuous = FALSE,location = NULL)
tsumbs <- c((trme[[1]])[,1], (trme[[2]])[,1], (trme[[3]])[,1])

return(tsumbs)
}

cl <- makeCluster(detectCores(), type='PSOCK')
registerDoParallel(cl)

set.seed(123456)
trutest <- replicate(1000, mybs1(truedata))

registerDoSEQ()

as.matrix(trutest)

maxpsu <- apply(trutest, 1, max)
minpsu <- apply(trutest,1, min)
meanpsu <- apply(trutest,1, mean)

lala <- cbind(meanpsu, minpsu, maxpsu)

trueme[[1]] <- cbind(trueme[[1]], lala[c(1:3),])
trueme[[2]] <- cbind(trueme[[2]], lala[c(4:6),])
trueme[[3]] <- cbind(trueme[[3]], lala[c(7:9),])

trumesum <- c((trueme[[1]])[,1], (trueme[[2]])[,1], (trueme[[3]])[,1])

lala2 <- as.data.frame(cbind(trumesum, lala))

signtrutest <- sign(trutest)
signtru <- sign(lala2$trumesum)

signchan <- (1- rowMeans((signtru - signtrutest)==0))

#Type1 errorate 1
cat_zsubchao_3cat <- sample(truedata$cat_zsubchao_3cat)

errdata <- cbind(truedata[,c("gmakerule_rec", "log_fsize", "avgofghhcomm_pcthh")],
                 cat_zsubchao_3cat)

erreg<-oglmx(model, data=errdata, link="logit", constantMEAN = FALSE,
             constantSD = FALSE,delta=0,threshparam = NULL)

```

```

errme <- margins.oglmx(treg, Vars = NULL, outcomes = "All",
                      atmeans = TRUE, AME = FALSE,
                      ascontinuous = FALSE, location = NULL)

perrme <- c((errme[[1]])[,4], (errme[[2]])[,4], (errme[[3]])[,4])

mybs2<- function(datx){
bsdata<- sample_n(datx, size=84,replace= FALSE)
rerrreg <- oglmx(model, data=bsdata, link="logit", constantMEAN = FALSE,
                 constantSD = FALSE,delta=0,threshparam = NULL)
rerrme <- margins.oglmx(rerrreg, Vars = NULL, outcomes = "All", AME = FALSE,
                       ascontinuous = FALSE,location = NULL)
rersumbs <- c((rerrme[[1]])[,4], (rerrme[[2]])[,4], (rerrme[[3]])[,4])

return(rersumbs)
}

library(doParallel)
cl <- makeCluster(detectCores(), type='PSOCK')
registerDoParallel(cl)

set.seed(123456)
rerrtest <- replicate(1000, mybs2(errdata))

type1errorrate.05 <- rowMeans(rerrtest<.05)
type1errorrate.1 <- rowMeans(rerrtest<.1)

#Type1errorrate 2 traditional permutation test

type12 <- function(datz){
  functiondata <- datz
  functiondata$shuffley <- sample(functiondata$cat_zsubchao_3cat)
  model3 <- as.ordered(shuffley) ~ log_fsize + as.factor(gmakerule_rec) + avgofghhcomm_pcthh
  rerrreg <- oglmx(model3, data=functiondata, link="logit", constantMEAN = FALSE,
                 constantSD = FALSE,delta=0,threshparam = NULL)
  rerrme <- margins.oglmx(rerrreg, Vars = NULL, outcomes = "All", AME = FALSE,
                       ascontinuous = FALSE,location = NULL)
  rersumbs <- c((rerrme[[1]])[,4], (rerrme[[2]])[,4], (rerrme[[3]])[,4])

  return(rersumbs)
}

permtest <- replicate(1000, type12(repdat))

type1permerrorrate.05 <- rowMeans(permtest<.05)
type1permerrorrate.1 <- rowMeans(permtest<.1)

errsummary <- cbind(signchan, type1errorrate.05, type1errorrate.1,
                   type1permerrorrate.05, type1permerrorrate.1)

```

```

colnames(errsummary) <- c("Change in Sign", "Simulation:\n Type 1 Error Rate (0.05)",
  "Simulation:\n Type 1 Error Rate (0.1)", "Permutation:\n Type 1
  Error Rate (0.05)", "Permutation:\n Type 1 Error Rate (0.5)")

rownames(errsummary) <- c("Rulemaking Participation", "Forest Size",
  "Commercial Livelihoods", "Rulemaking Participation",
  "Forest Size", "Commercial Livelihoods",
  "Rulemaking Participation", "Forest Size",
  "Commercial Livelihoods")

tableerll <- xtable(errsummary[c(1:3),], align = "ccccc", latex.environments = "center",
  caption="Joint Outcome Category (Low-Low): Marginal effects of ordered Logit Regress",
  digits=c(0,3,3,3,3,3))

align(tableerll) <- c( 'l', 'p{.80in}', 'p{.80in}', 'p{.8in}', 'p{.8in}', 'p{.8in}' )

errll <- print(tableerll, comment = FALSE)

tableerto <- xtable(errsummary[c(4:6),], align = "ccccc", latex.environments = "center",
  caption="Joint Outcome Category (Trade-off): Marginal effects of ordered Logit Regress",
  digits=c(0,3,3,3,3,3))

align(tableerto) <- c( 'l', 'p{.80in}', 'p{.80in}', 'p{.8in}', 'p{.8in}', 'p{.8in}' )

errto <- print(tableerto)

tableerhh <- xtable(errsummary[c(7:9),], align = "ccccc", latex.environments = "center",
  caption="Joint Outcome Category (High-High): Marginal effects of ordered Logit Regress",
  digits=c(0,3,3,3,3,3))

align(tableerhh) <- c( 'l', 'p{.80in}', 'p{.80in}', 'p{.8in}', 'p{.8in}', 'p{.8in}' )

errhh <- print(tableerto, comment = FALSE)

```

Chao, Anne, Robin L. Chazdon, Robert K. Colwell, and Tsung-Jen Shen. 2005. "A New Statistical Approach for Assessing Similarity of Species Composition with Incidence and Abundance Data." *Ecology Letters* 8 (2): 148–59.

Epstein, Graham, Jessica Vogt, Michael Cox, and Luke Shimek. 2014. "Confronting Problems of Method in the Study of Sustainability." *Forest Policy and Economics* 42: 42–50.

Ostrom, Elinor, and Mary Beth Wertime. 1998. "IFRI Research Strategy." *Forest Resources and Institutions, FAO, Rome*, 217.

Persha, Lauren, Arun Agrawal, and Ashwini Chhatre. 2011. "Social and Ecological Synergy: Local Rulemaking, Forest Livelihoods, and Biodiversity Conservation." *Science* 331 (6024): 1606–8.

Poteete, Amy R., Marco A. Janssen, and Elinor Ostrom. 2010. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice*. Princeton University Press.