National College of Ireland

Configuration Manual

MSc Research Project
M.sc in Fintech

Adeniyi Peter Makinde
X18133681

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Adeniyi Peter Makinde
Student ID:	X18133681
Programme:	M.sc in Fintech **Year:** 2019
Module:	Msc Research Project
Lecturer:	Noel Cosgrave
Submission Due Date:	12/08/2019
Project Title:	Forecasting Economic Recession in Selected African Countries using Machine Learning Algorithms
Word Count:	…… **Page Count:** ……

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……

Date: 12th September 2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Forecasting Economic Recession in Selected African Countries using Machine Learning Algorithms

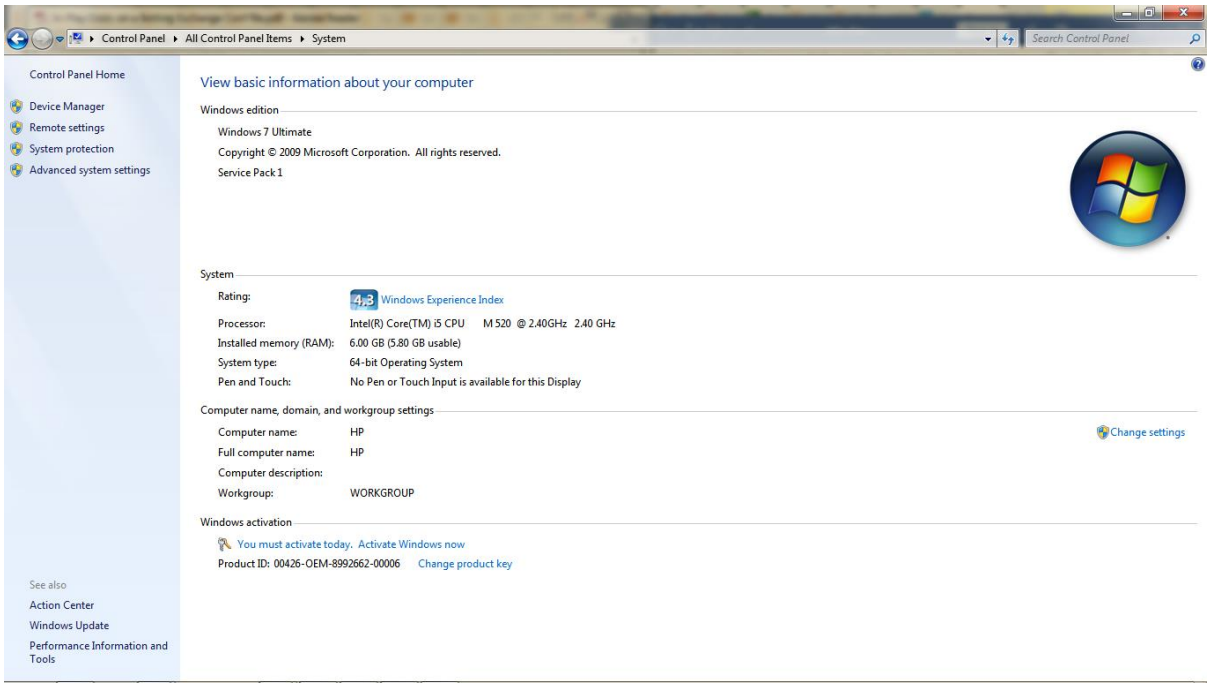
Configuration Manual

Adeniyi Peter Makinde
X18133681

1 R System Requirements

- Downloaded from: https://cran.r-project.org/mirrors.html
- Setup File: “R-3.2.2-win.exe” (65MB)
- *Testing is primarily on versions of Windows currently supported by Microsoft, mainly 64-bit Windows 7 and Server 2008
- Installation requires 150MB of disk space
- Memory Limit: We only test on versions of Windows currently supported by Microsoft, mainly 64-bit Windows 7 and Server 2008
- Further Requirements see: http://ftp.heanet.ie/mirrors/cran.r-project.org/

2 Personal system Requirement



Software run on:

- Manufacturer: Hewlett Packard
- Model: HP Elitebook
- Processor: intel® Core™ i5
- Ram Memory: 6.00GB
- System type: 64-Bit Operating System
- HDD: 297GB

Data Source:

- Penn world productivity data
  Website: www.ggdc.net/pwt
- Bank of Canada commodity Indices to compare and evaluate possible contagion from this index2
  Website: www.bankofcanada.ca/rates/price-indexes/bcoi
- world bank Gdp growth datase
  Website: api.worldbank.org

3 R code:

```
install.packages("mctest")
install.packages("ppcor")
install.packages("corpcor")
install.packages("glmnet")
install.packages("mlbench")
install.packages("psych")
install.packages("stargazer")
install.packages("ROCR")
install.packages("naivebayes")
```

```
library(haven)
library(mctest)
library(ppcor)
library(GGally)
library(corpcor)
library(caret)
library(glmnet)
library(mlbench)
library(psych)
library(dplyr)
library(stargazer)
library(nnet)
library(ROCR)
library(naivebayes)
library(randomForest)
```

```
African<-read_csv("C:/Users/NEW USER/Desktop/Final Project 2019/Recession prediction
dataset/African.csv")
View(African)
```

```
str(African)
summary(African)


#Proportion

table(African$growthbucket)/sum(table(African$growthbucket))

#Test for Multicollinearity

Y<-African[c(50)]
X<-African[c(-50)]
cov<- cor2pcor(cov(X))
View(cov)
omcdiag(X,Y)
imcdiag(X,Y)


lg<-glm(growthbucket~.,African,family = "binomial")
summary(lg)

# Backward Stepwise Regression

step(lg,direction = "backward")->bw

lgt<-glm(growthbucket ~ emp + emp_to_pop_ratio + cda + cn + ck + ctfp +
   rkna + delta + pl_da + csh_c + csh_r + pl_g +
  pl_x + pl_n + fish + total_change + energy_change,African,family = "binomial")

summary(lgt)

##New Dataset

data<-African[c("growthbucket","emp","emp_to_pop_ratio","cda",
            "cn","ck","ctfp","pl_x","pl_n","fish","total_change","energy_change")]
View(data)
data$growthbucket<-as.factor(data$growthbucket)
summary(data$growthbucket)

##DQR

DQR<-function(df){
  n<-sapply(df,function(x){is.numeric(x)})
  df_numerics<-df[,n]
  instances<-sapply(df_numerics,FUN = function(x){length(x)})
  missing<-sapply(df_numerics,FUN = function(x){sum(is.na(x))})
  missing<-missing/instances*100
  unique<-sapply(df_numerics,FUN = function(x){length(unique(x))})
  quantiles<-t(sapply(df_numerics,FUN =  function(x){quantile(x)}))
  means<-sapply(df_numerics,FUN =  function(x){mean(x)})
  sds<-sapply(df_numerics,FUN = function(x){sd(x)})

  df_numerics<-data.frame(Feature=names(df_numerics),
                          Instances=instances,
                          Missing=missing,
                          Cardinality=unique,
                          Min=quantiles[,1],
```

```
                    FirstQuantile=quantiles[,2],
                    Median=quantiles[,3],
                    ThirdQuantile=quantiles[,4],
                    Max=quantiles[,5],
                    Mean=means,
                    Stdev=sds)

  rownames(df_numerics)<-NULL
  return(df_numerics)
}
df_numerics<-DQR(data)
df_numerics

##Descriptive Statistics

summary(data)
table(data$growthbucket)
table(data$growthbucket)/sum(table(data$growthbucket))

#Logit Model Coeficients

logit<-glm(growthbucket~.,data = data,family = binomial(link = "logit"))
summary(logit)

#Logit model odds ratio

exp(logit$coefficients)

#Probit Model Coefficients

probit<-glm(growthbucket~.,data = data,family = binomial(link = "probit"))
summary(probit)

#logit model average marginal effects

logitScalar<-mean(dlogis(predict(logit,type="link")))
logitScalar<-coef(logit)

#Probit model average marginal effects

probitScalar<-mean(dnorm(predict(probit,type="link")))
probitScalar<-coef(probit)

#Logit Model Predicted Probabilities

plogit<-predict(logit,type = "response")
summary(plogit)

#Probit Model predicted Probabilities

pprobit<-predict(probit,type="response")
summary(pprobit)

#Percent correctly predicted values

table(true=data$growthbucket,pred=round(fitted(probit)))->acc1
table(true=data$growthbucket,pred=round(fitted(logit)))->acc2
```

```
#MaFadden's Pseudo R-Square

probit0<-update(probit,formula=data$growthbucket~1)
McFadden<-1-as.vector(logLik(probit)/logLik(probit0))
McFadden

#Misclassification
##Logit

sum(diag(acc2))/sum(acc2)
1-sum(diag(acc2))/sum(acc2)

##Probit

sum(diag(acc1))/sum(acc1)
1-sum(diag(acc1))/sum(acc1)


#GOF-test

with(logit,pchisq(null.deviance-deviance,df.null-df.residual,lower.tail=F))

### NAIVE BAYES ALGORITHM

Bayes<-naive_bayes(growthbucket~.,data = data)
summary(Bayes)
plot(Bayes)

##Prediction

pred_bayes<-predict(Bayes,data,type = "prob")
tab_bayes<-(cbind(pred_bayes,Bayes))


#Confusion Matrix

pred1_bayes<-predict(Bayes,data)
confusion_matrix<-table(pred1_bayes,data$growthbucket)
 print(confusion_matrix)

#Misclassification
 sum(diag(confusion_matrix))/sum(confusion_matrix)
1-sum(diag(confusion_matrix))/sum(confusion_matrix)

##Random Forest

Forest<-randomForest(growthbucket ~., data = data)
Forest

##Prediction

pred_forest<-predict(Forest,data, type = "class")

#### checking classification accuracy

table(pred_forest,data$growthbucket)->conf_forest
```

```
print(conf_forest)
mean(pred_forest==data$growthbucket)

1-sum(diag(table(conf_forest)))/sum(table(conf_forest))

#Variables of Importance

importance(Forest)
varImpPlot(Forest)
```

plots

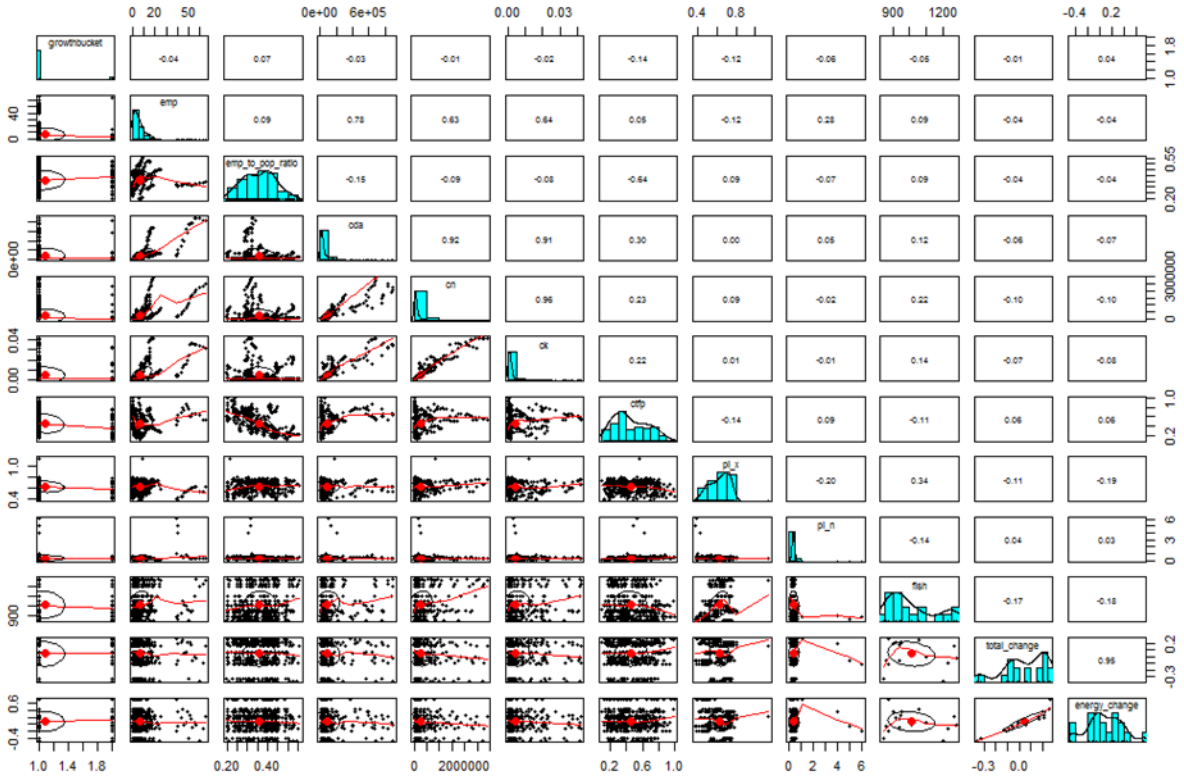
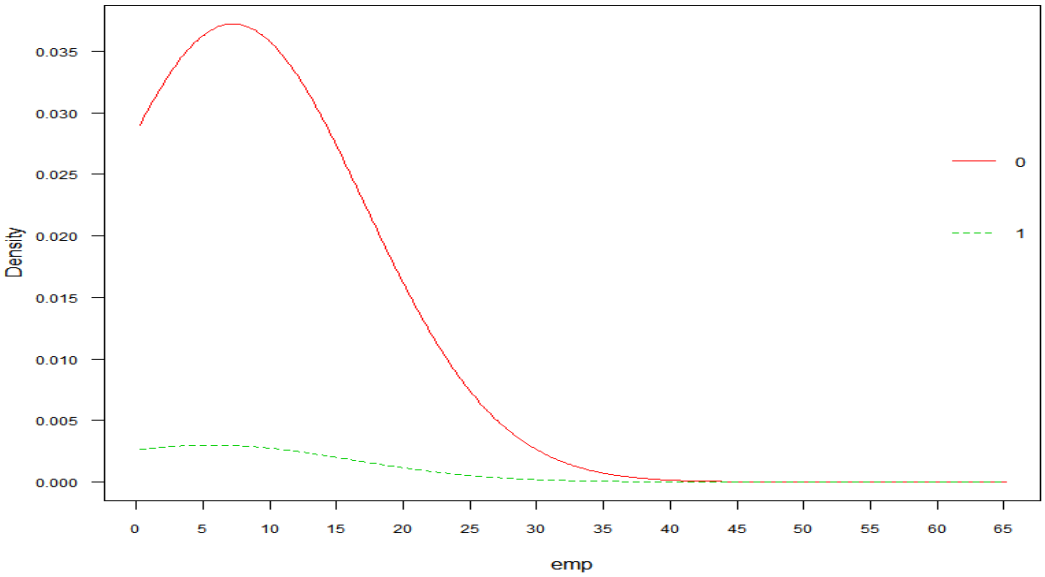
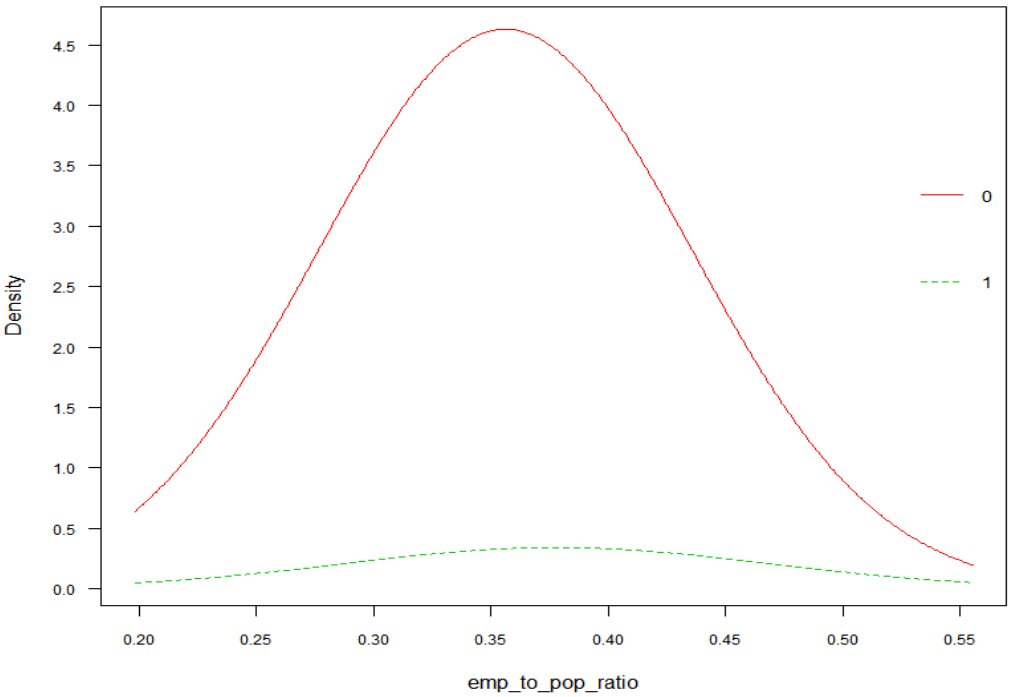
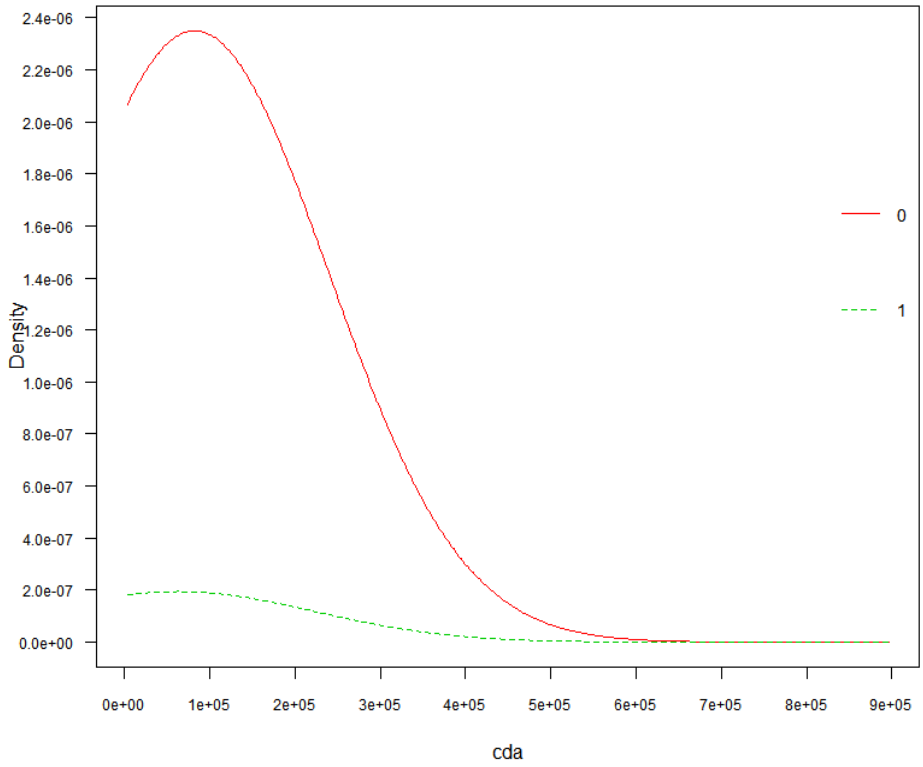
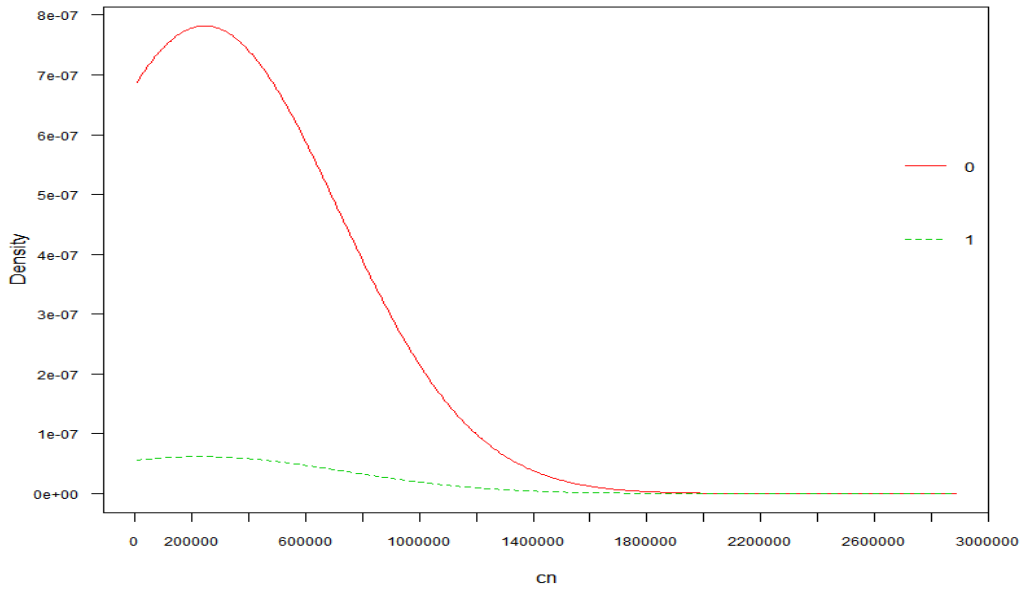


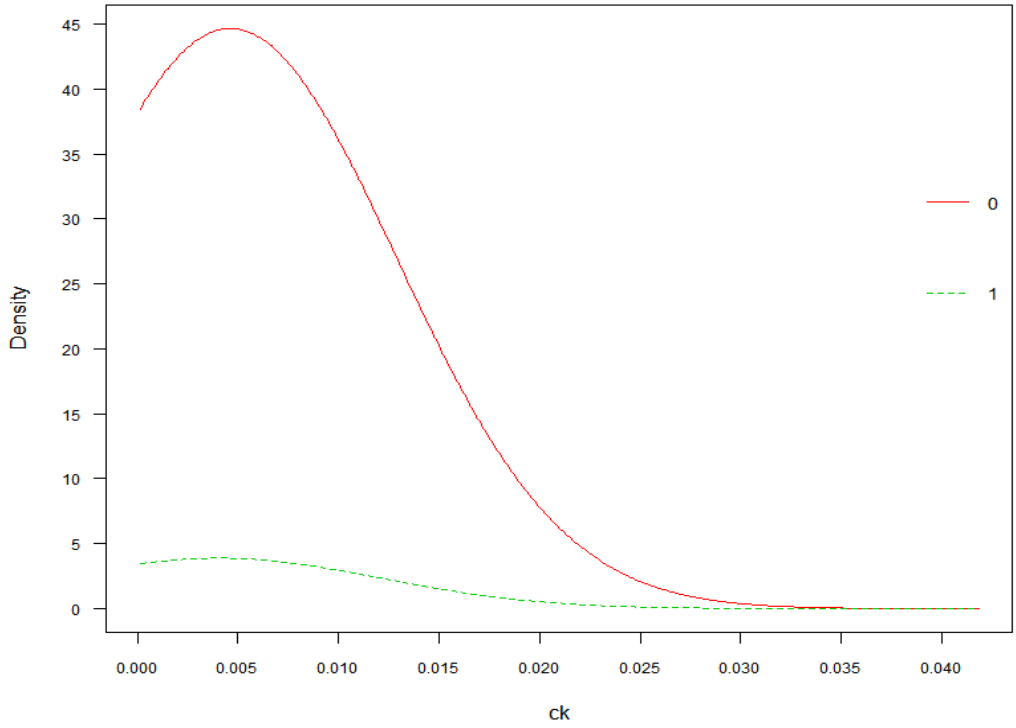
Figure 1: Correlation Matrix

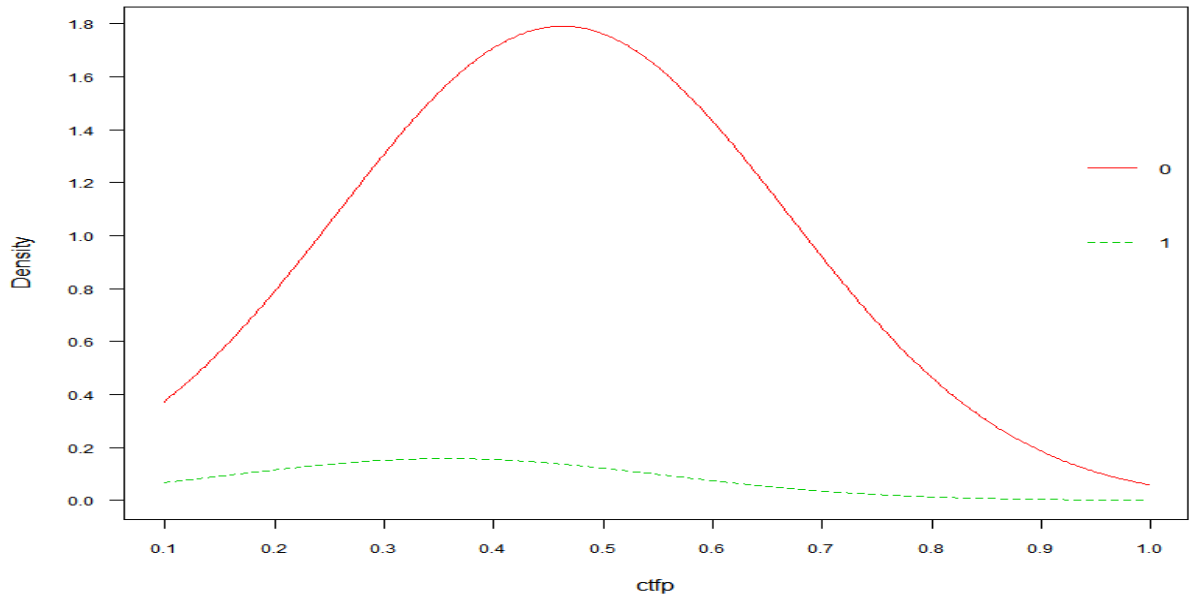

Density
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
5
10
15
20
25
30
35
40
45
50
55
60
65
emp
0
1

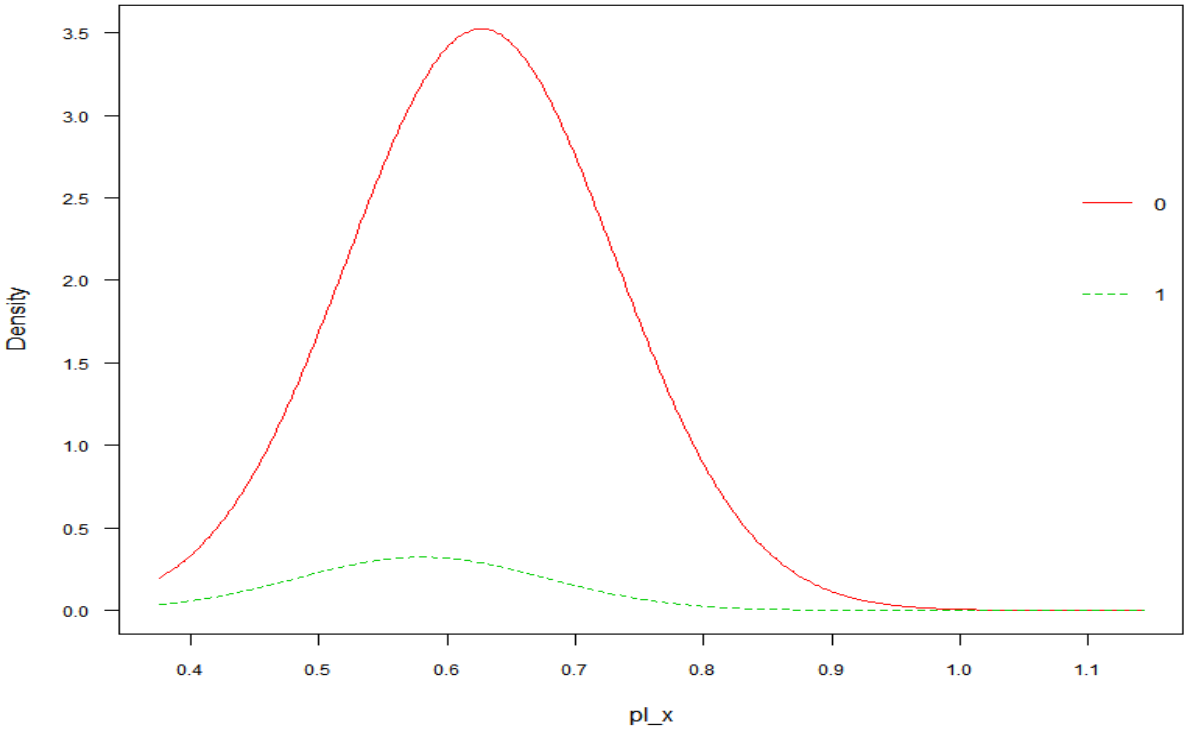

Density
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
emp_to_pop_ratio
0
1

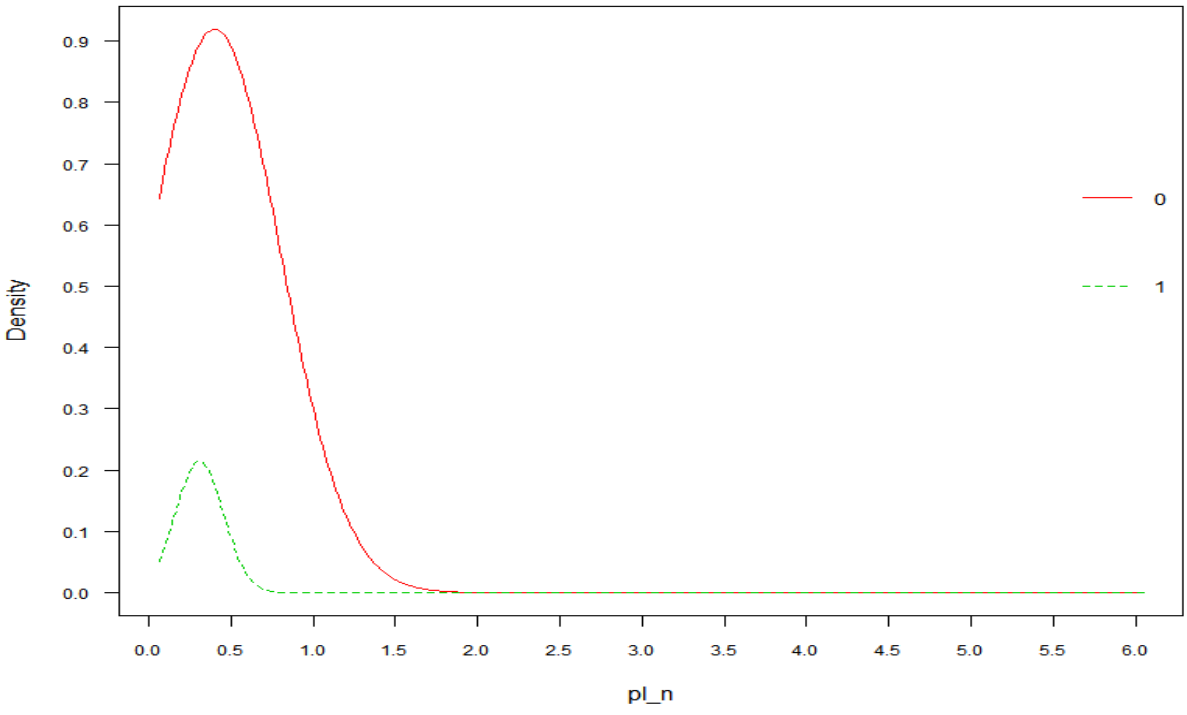

2.4e-06
2.2e-06
2.0e-06
1.8e-06
1.6e-06
1.4e-06
1.2e-06
1.0e-06
8.0e-07
6.0e-07
4.0e-07
2.0e-07
0.0e+00
Density
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
7e+05
8e+05
9e+05
cda
0
1

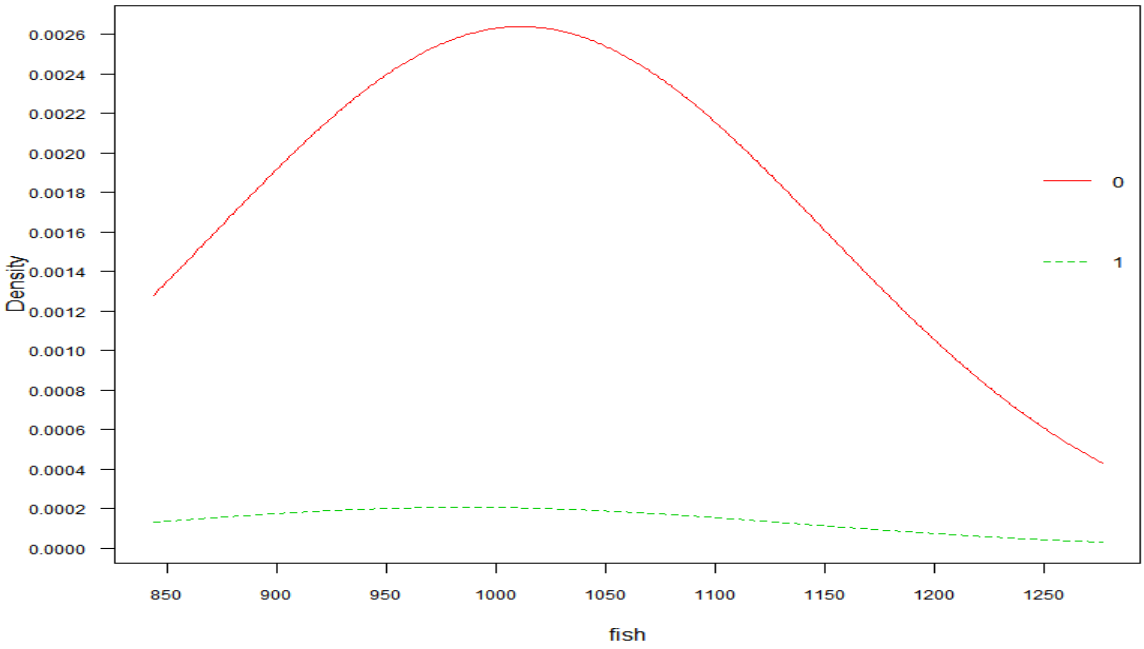

8e-07
7e-07
6e-07
5e-07
4e-07
3e-07
2e-07
1e-07
0e+00
Density
0
200000
600000
1000000
1400000
1800000
2200000
2600000
3000000
cn
0
1

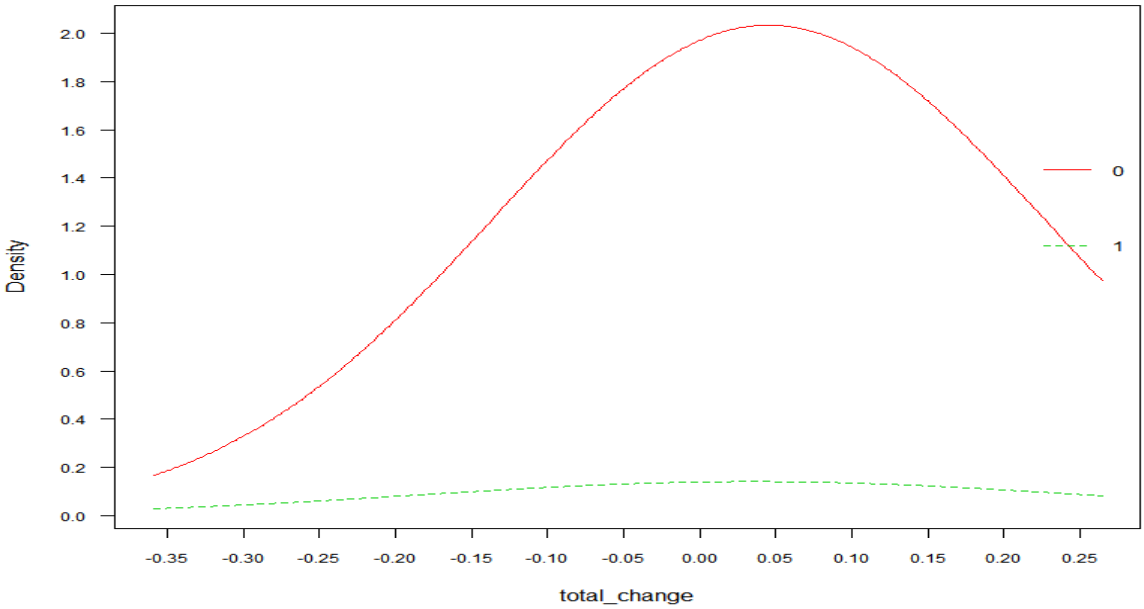

Density
45
40
35
30
25
20
15
10
5
0
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0.035
0.040
ck
0
1

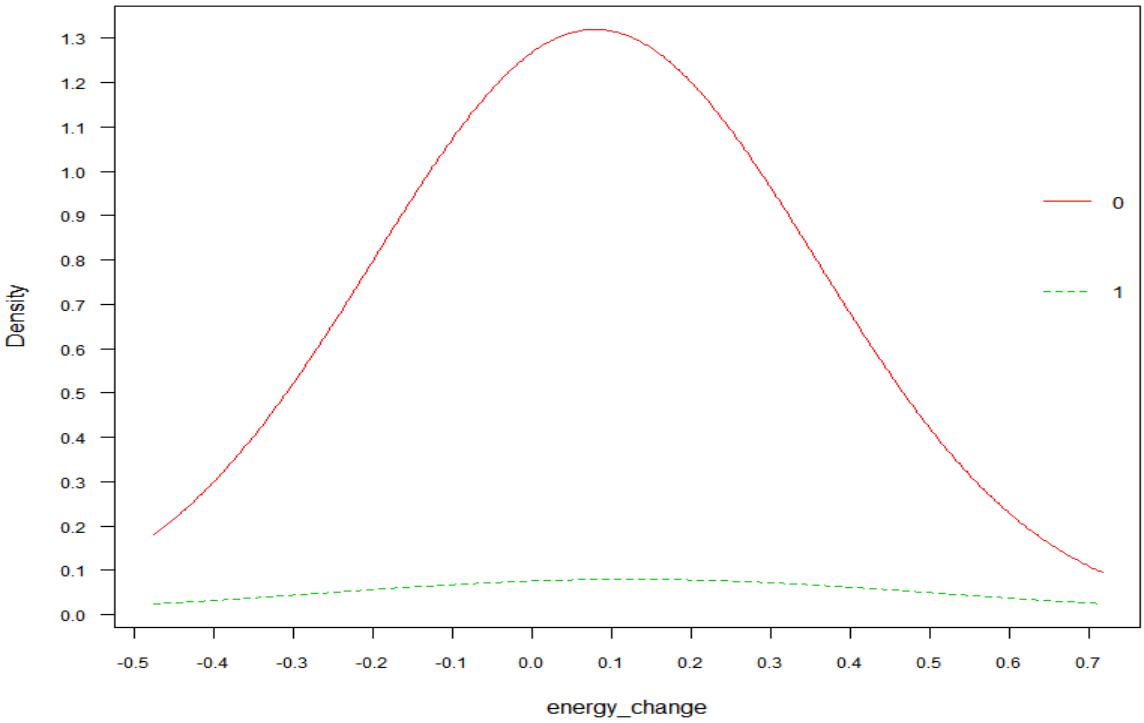

Density
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
ctfp
0
1

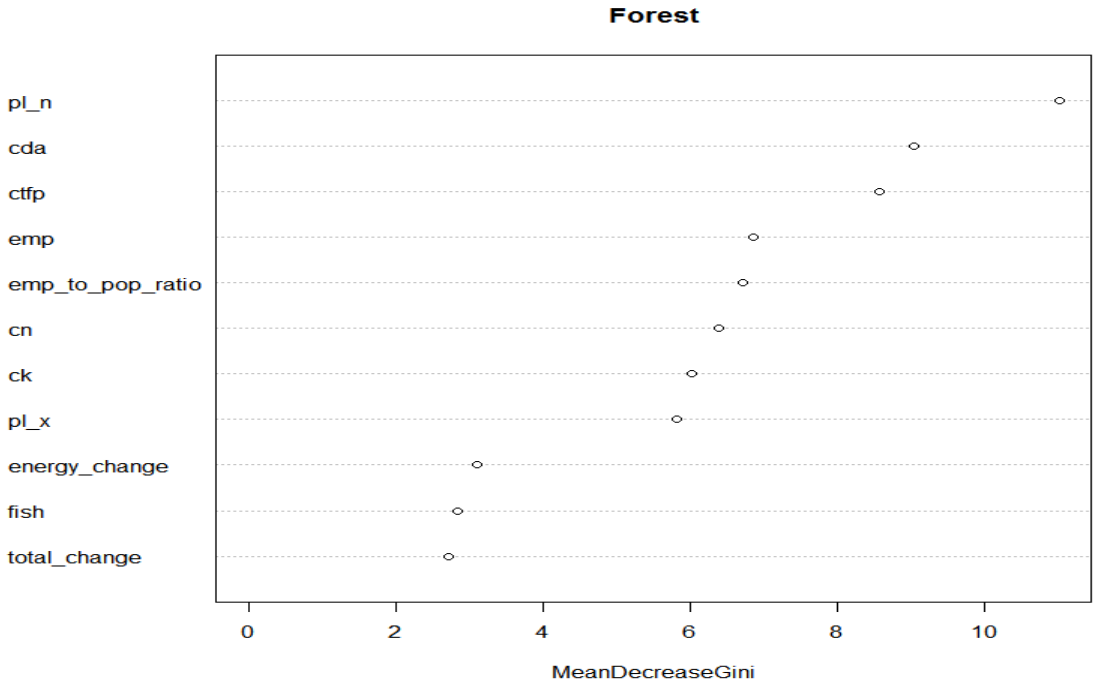

Density
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
pl_x
0
1


Density
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
pl_n
0
1


Density
0.0026
0.0024
0.0022
0.0020
0.0018
0.0016
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
850
900
950
1000
1050
1100
1150
1200
1250
fish
0
1


Density
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.35
-0.30
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
total_change
0
1


Density
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
energy_change
0
1


Forest
pl_n
cda
ctfp
emp
emp_to_pop_ratio
cn
ck
pl_x
energy_change
fish
total_change
0
2
4
6
8
10
MeanDecreaseGini