natural-disasters.R

matteoanzano

2023-11-16

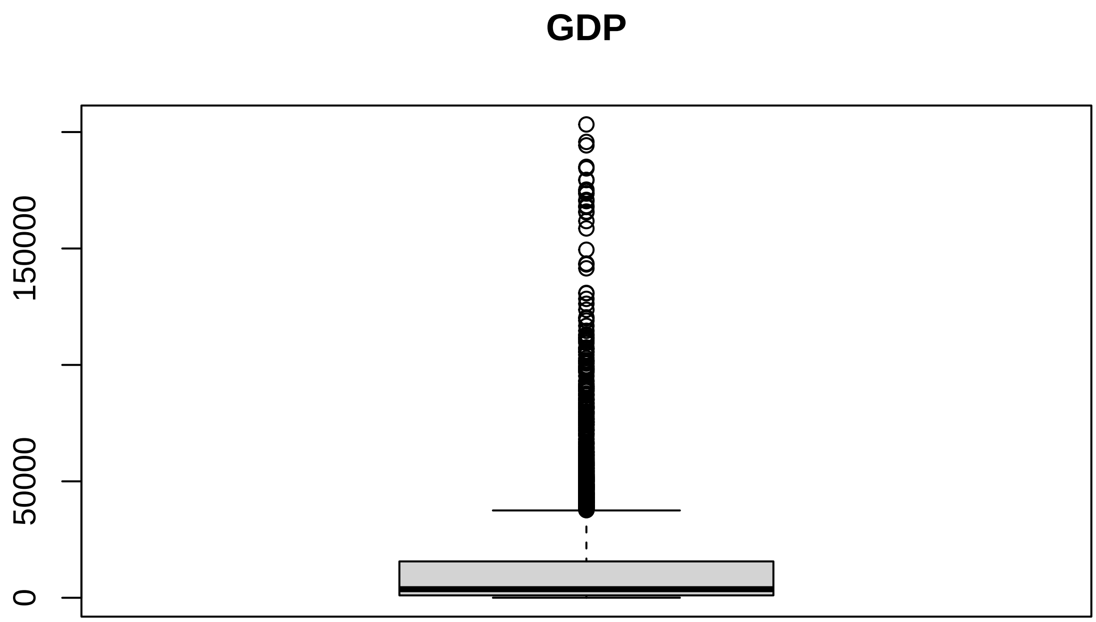
```r
library(readxl)
library(mlVAR)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

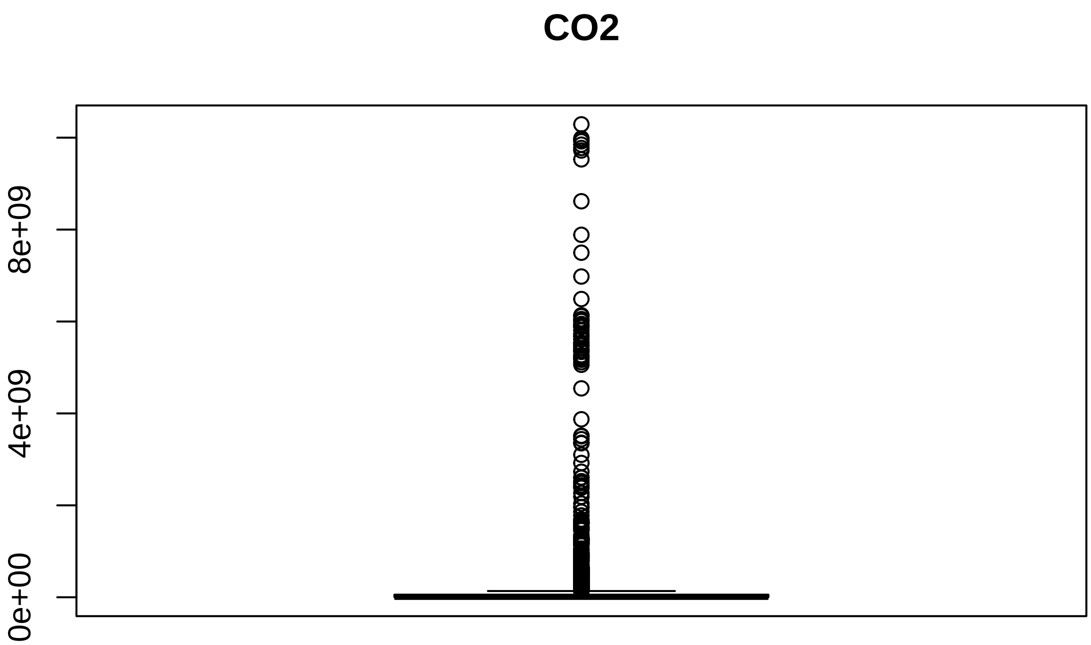
```r
library(ggpubr)
library(plm)
```

```
##
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

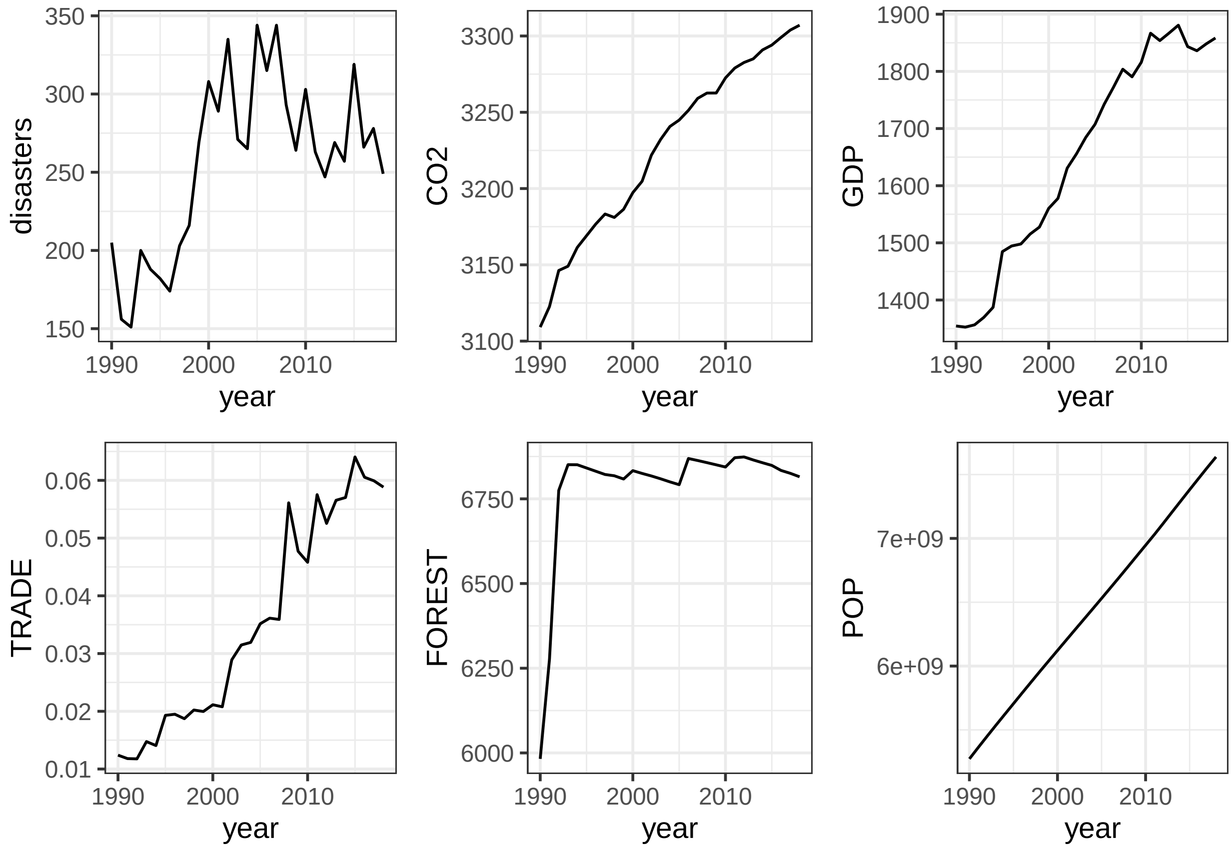
```r
rm(list=ls())
dataset=read.csv("datasig.csv")
dataset=dataset[,-1]
dataset$co2=dataset$Annual.CO2.emissions
dataset$gdp=dataset$GDP..current.US../dataset$Population..total
dataset$trade=dataset$Trade....of.GDP./dataset$Population..total
dataset$pop=dataset$Population..total
dataset$forest=dataset$Forest.area....of.land.area.
dataset$mob=dataset$Mobile.cellular.subscriptions..per.100.people.
dataset=dataset%>%dplyr::select(country,year,disasters,co2,gdp,trade,forest, pop)


boxplot(dataset$gdp, main="GDP")
```



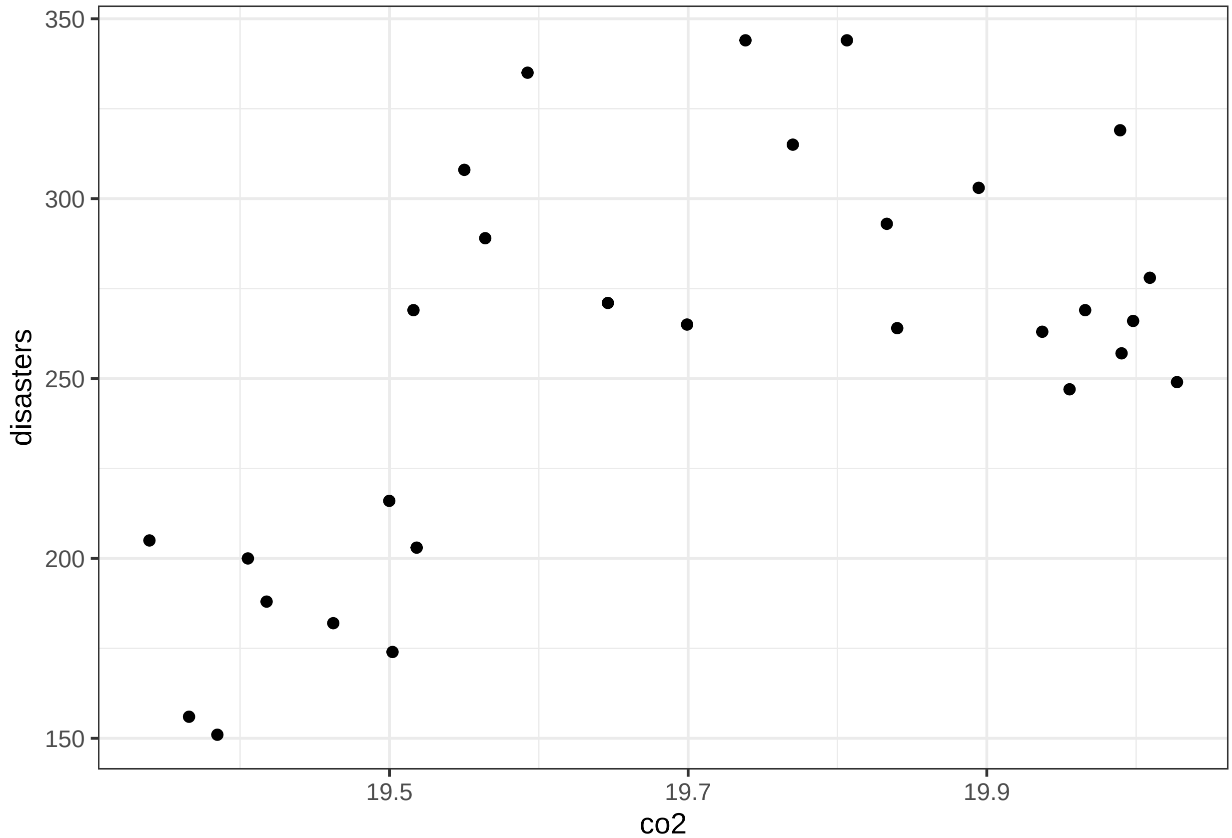

```
boxplot(dataset$co2, main="CO2")
```




```r
dataset$gdp=log(dataset$gdp)
dataset$co2=log(dataset$co2)
# come sono distribuiti i disastri?
g1=dataset%>%group_by(year)%>%summarize(disasters=sum(disasters, na.rm=T))%>%ggplot()+
  geom_line(aes(y=disasters, x=year))+
  theme_bw()
g2=dataset%>%group_by(year)%>%summarize(CO2=sum(co2, na.rm=T))%>%ggplot()+
  geom_line(aes(y=CO2, x=year))+
  theme_bw()
g3=dataset%>%group_by(year)%>%summarize(GDP=sum(gdp, na.rm=T))%>%ggplot()+
  geom_line(aes(y=GDP, x=year))+
  theme_bw()
g4=dataset%>%group_by(year)%>%summarize(TRADE=sum(trade, na.rm=T))%>%ggplot()+
  geom_line(aes(y=TRADE, x=year))+
  theme_bw()
g5=dataset%>%group_by(year)%>%summarize(FOREST=sum(forest, na.rm=T))%>%ggplot()+
  geom_line(aes(y=FOREST, x=year))+
  theme_bw()
g6=dataset%>%group_by(year)%>%summarize(POP=sum(pop))%>%ggplot()+
  geom_line(aes(y=POP, x=year))+
  theme_bw()
ggarrange(g1,g2,g3,g4,g5,g6)
```



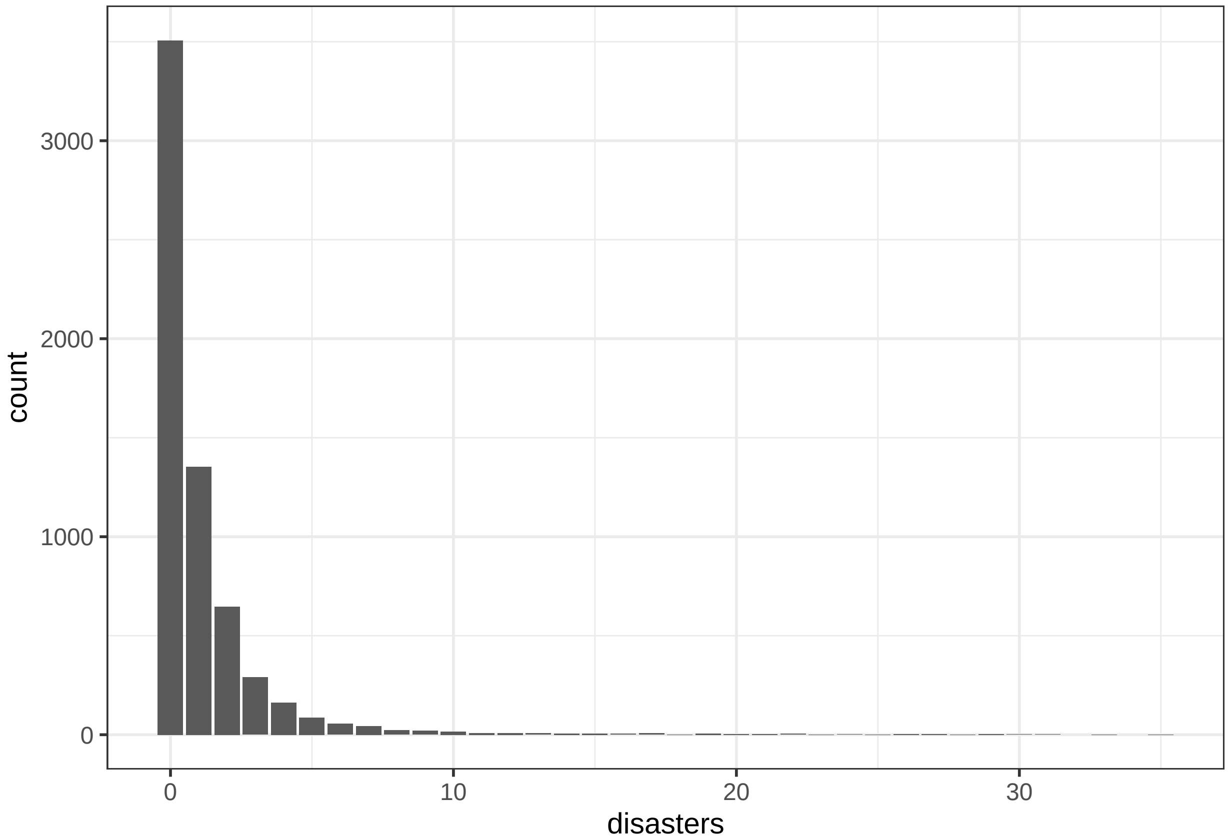

```
# nostro scopo è comprendere se c'è dipendenza del numero di disastri dalle emissioni di co2.
#E' di interesse cioè comprendere se un aumento delle emissioni di co2, a parità di gdp, popolazioen
# e eventuali altre variabili di controllo comporta un aumento del numero di disastri.

dataset%>%group_by(year)%>%summarize(disasters=sum(disasters, na.rm=T), co2=sum(co2*pop, na.rm=T)/sum(po
  geom_point(aes(y=disasters, x=co2))+
  theme_bw()
```




```
table(dataset$year)
```

```
##
## 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005
##  217  217  217  217  217  217  217  217  217  217  217  217  217  217  217  217
## 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
##  217  217  217  217  217  217  217  217  217  217  217  217  217
```

```
ggplot(dataset)+
  geom_bar(aes(x=disasters))+
  theme_bw()
```




```r
# poichè c'è un numero molto elevato di casi con 0 disastri naturali, è di interesse costruire modelli
# che permettono di capire quali sono i fattori per cui si verifica un disastro naturale. La tratto
# quindi come variabile dummy, che vale 1 se si è verificato almeno un disastro, 0 altrimenti

dataset$disaster=ifelse(dataset$disasters>0,1,0)
data=dataset%>%dplyr::select(-disasters)
# modellistica#
m1<-glm(disaster~co2, data=data, family="binomial")
summary(m1)
```

```
##
## Call:
## glm(formula = disaster ~ co2, family = "binomial", data = data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.3827     0.1841  -23.80   <2e-16 ***
## co2           0.2666     0.0114   23.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 8144.6  on 5894  degrees of freedom
## Residual deviance: 7524.7  on 5893  degrees of freedom
```

```
##   (398 observations deleted due to missingness)
## AIC: 7528.7
##
## Number of Fisher Scoring iterations: 4
```

```
exp(0.2419)-1
```

```
## [1] 0.2736668
```

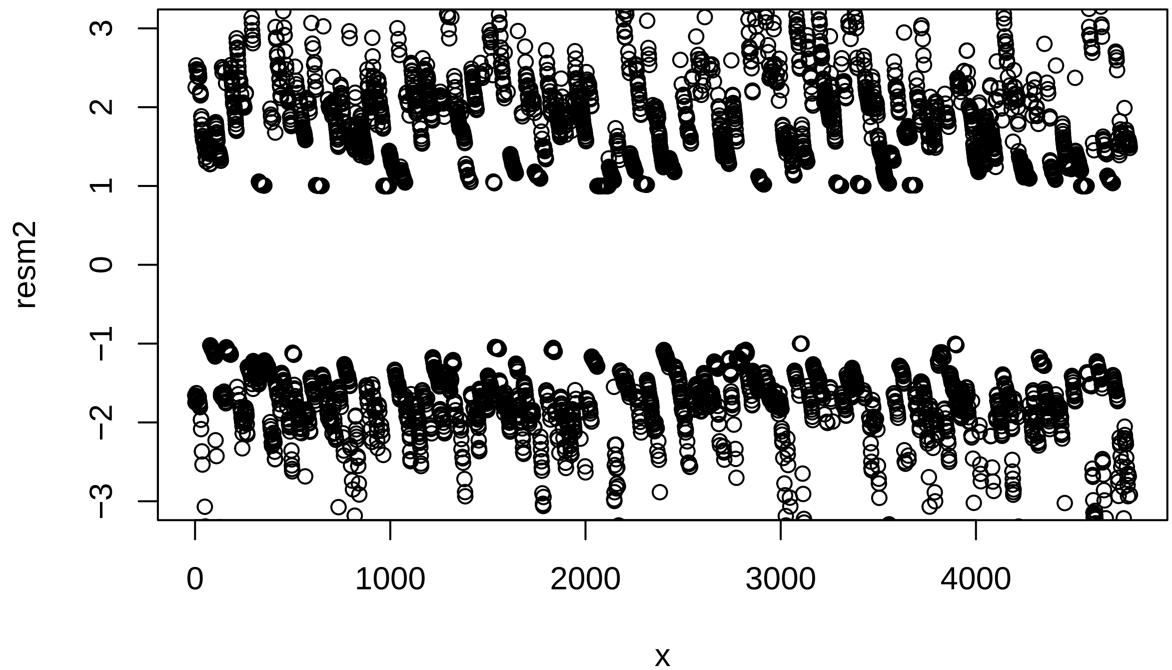
```
# Tenendo conto della sola associazione tra disastri e co2, l'effetto sembra essere positivo.
#In particolare l'aumento delle emissioni di una tonnellata comporta un aumento delle emissioni del 27

m2<-glm(disaster~co2+gdp+trade+forest+pop+year, data=data, family="binomial")
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

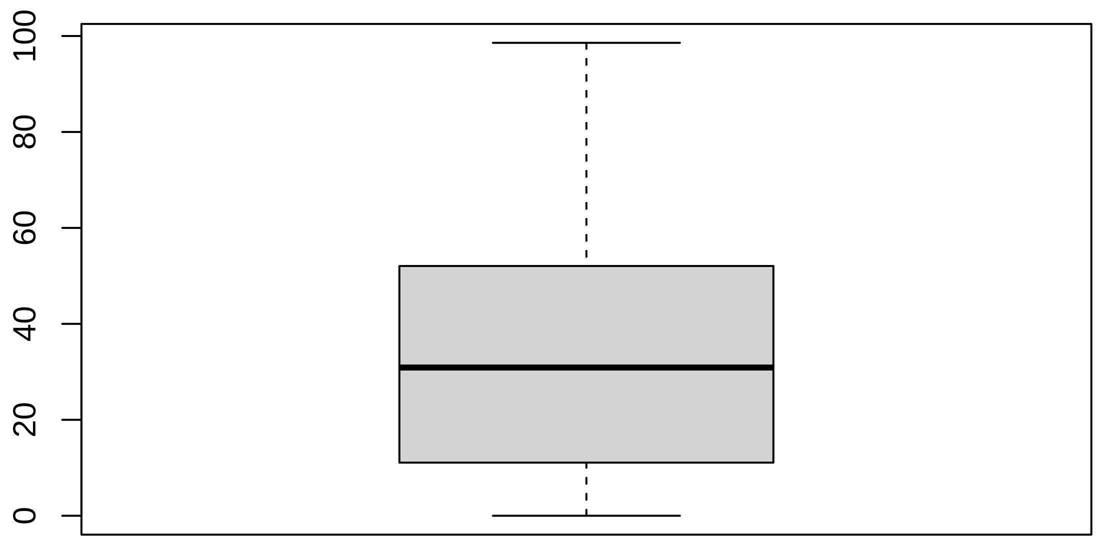
```
summary(m2)
```

```
##
## Call:
## glm(formula = disaster ~ co2 + gdp + trade + forest + pop + year,
##     family = "binomial", data = data)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.368e+01  8.498e+00  -6.317 2.67e-10 ***
## co2          9.281e-02  2.902e-02   3.198  0.00138 **
## gdp         -2.057e-01  3.347e-02  -6.147 7.91e-10 ***
## trade       -7.042e+02  1.511e+02  -4.660 3.16e-06 ***
## forest       9.488e-03  1.408e-03   6.737 1.61e-11 ***
## pop          3.064e-08  2.556e-09  11.988  < 2e-16 ***
## year         2.652e-02  4.234e-03   6.265 3.74e-10 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6633.6  on 4788  degrees of freedom
## Residual deviance: 5521.2  on 4782  degrees of freedom
##   (1504 observations deleted due to missingness)
## AIC: 5535.2
##
## Number of Fisher Scoring iterations: 7
```

```
# Tenendo però conto anche delle restanti variabili, il suo effetto sembra invece essere negativo.
resm2=m2$residuals
x=1:length(resm2)
plot(resm2~x, ylim=c(-3,3))
```




```
# i residui sono ovviamente non casuali. Questo è dovuto al fatto che sono dati di tipo panel,
# cioè abbiamo osservazioni ripetute per lo stesso individuo nel tempo

boxplot(dataset$forest)
```




```
library(lme4)
```

```
## Loading required package: Matrix
```

```
lmm1=lmer(disaster~co2+gdp+trade+forest+as.numeric(year)+I(as.numeric(year)^2)+(1|country), data=data)
```

```
## Warning: Some predictor variables are on very different scales: consider
## rescaling
```

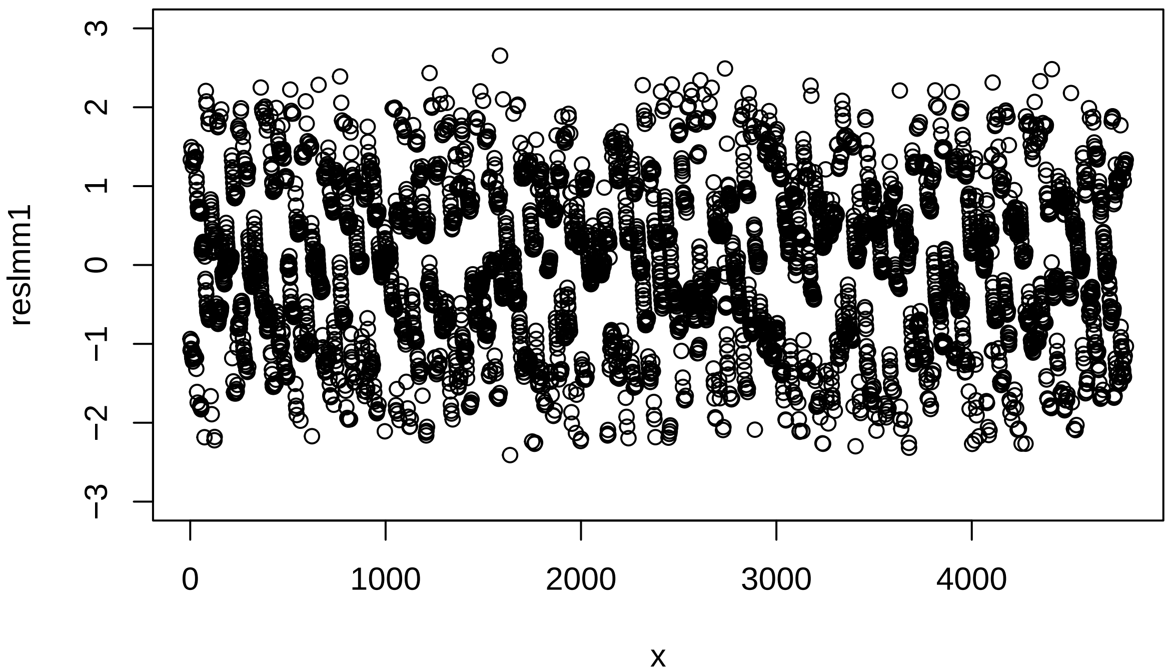
```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00486574 (tol = 0.002, component 1)
```

```
summary(lmm1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## disaster ~ co2 + gdp + trade + forest + as.numeric(year) + I(as.numeric(year)^2) +
##     (1 | country)
##    Data: data
##
## REML criterion at convergence: 5315.3
##
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max 
## -2.40987 -0.72949  0.01107  0.72589  2.65432 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  country  (Intercept) 0.04711  0.2170  
##  Residual             0.16203  0.4025  
## Number of obs: 4789, groups:  country, 186
## 
## Fixed effects:
##                         Estimate Std. Error t value
## (Intercept)           -2.492e+03  3.853e+02  -6.470
## co2                    7.759e-02  7.646e-03  10.148
## gdp                   -8.687e-02  1.076e-02  -8.076
## trade                 -1.379e+00  1.641e+01  -0.084
## forest                 1.809e-03  7.038e-04   2.571
## as.numeric(year)       2.479e+00  3.844e-01   6.448
## I(as.numeric(year)^2) -6.164e-04  9.590e-05  -6.427
## 
## Correlation of Fixed Effects:
##             (Intr) co2    gdp    trade  forest as.n()
## co2          0.027                                   
## gdp         -0.020 -0.421                            
## trade        0.000  0.359 -0.231                     
## forest       0.013  0.150 -0.074  0.133              
## as.nmrc(yr) -1.000 -0.027  0.021  0.000 -0.013       
## I(s.nm()^2)  1.000  0.027 -0.023  0.000  0.013 -1.000
## fit warnings:
## Some predictor variables are on very different scales: consider rescaling
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00486574 (tol = 0.002, component 1)
```

```
reslmm1=summary(lmm1)[["residuals"]]
x=1:length(reslmm1)
plot(reslmm1~x, ylim=c(-3,3))
```




```r
# ora i residui sembrano casuali, quindi un modello di questo tipo è più adatto



# modellazione VAR####
data7=data%>%dplyr::select(disaster,co2,gdp,forest,trade, country)
mvar7=mlVAR(data7, colnames(data7)[-6], colnames(data7)[6], lags=1)
```

```
## 'estimator' argument set to 'lmer'

## 'temporal' argument set to 'correlated'

## 'contemporaneous' argument set to 'correlated'

## Warning in mlVAR(data7, colnames(data7)[-6], colnames(data7)[6], lags = 1): 27
## subjects detected with < 20 measurements. This is not recommended, as
## within-person centering with too few observations per subject will lead to
## biased estimates (most notably: negative self-loops).

## Estimating temporal and between-subjects effects

##   |                                                                              |

## Estimating contemporaneous effects
```

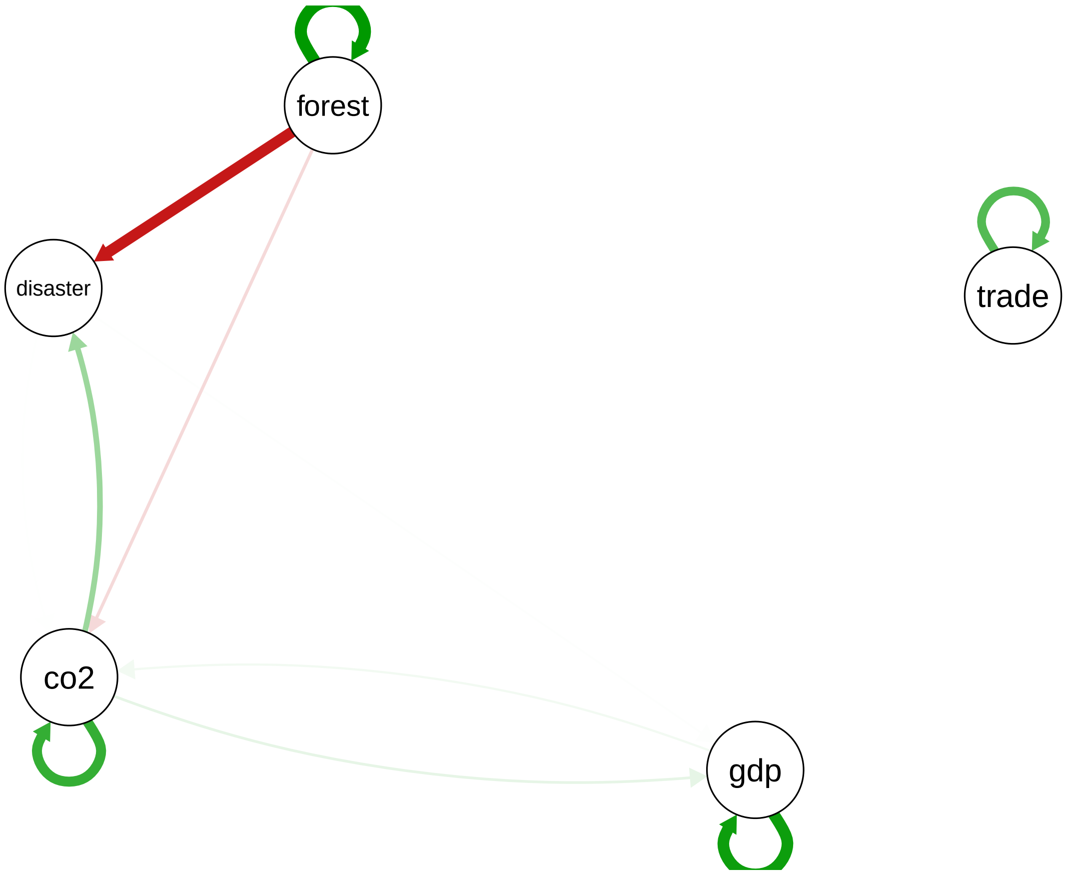
```
##   |                                                                              |
```

```
## Computing random effects
```

```
##   |                                                                              |
```

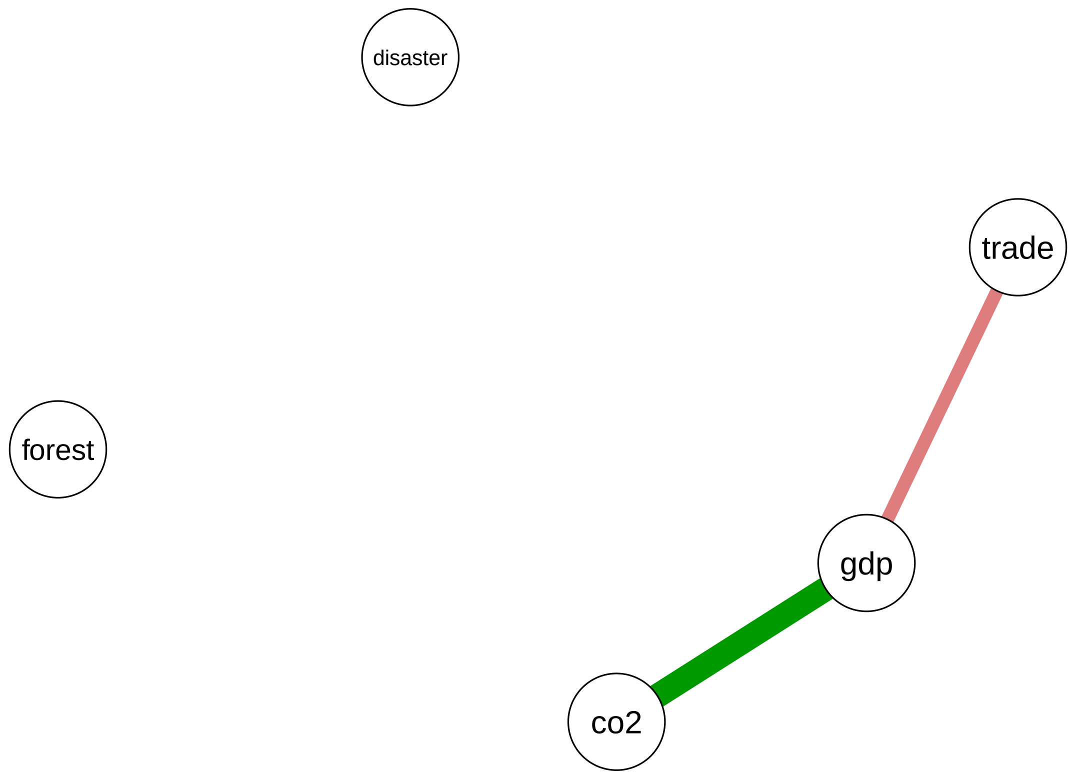
```r
plot(mvar7, "temporal")
```

```
## 'nonsig' argument set to: 'hide'
```



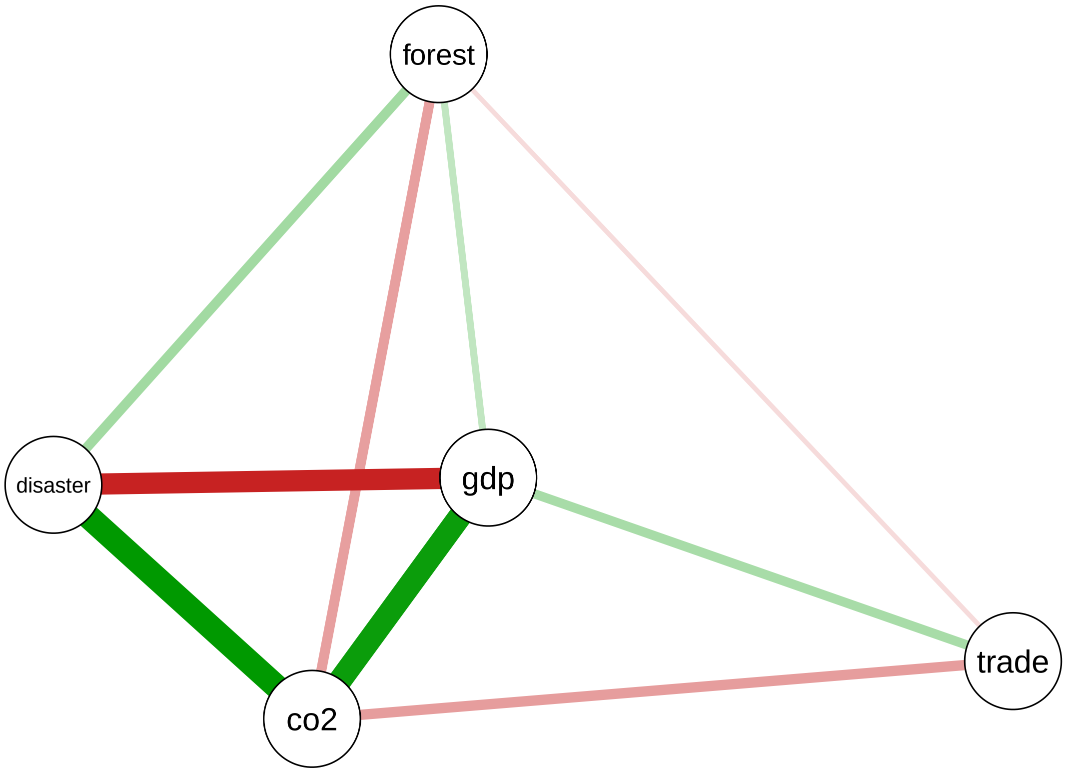

```r
plot(mvar7, "contemporaneous")
```

```
## 'nonsig' argument set to: 'hide'
```




```
plot(mvar7, "between")
```

```
## 'nonsig' argument set to: 'hide'
```




```
# Il modello utilizzato è di interesse perchè permette di comprendere vari tipi di relaizoni

# La relazione di maggiore interesse è quella tra la variabile emissioni di CO2 e i disastri naturali
# E' una relazione temporale positiva, il che indica che un incremento nelle emissioni comporta un incr
# nella probability di avere disastri naturali.
```