Quantile Based GAM

2023-07-05

In this notebook, we try to run quantile-based GAM

We run regression on quantiles, instead than on the mean. We start from the median

```
final_model_females_qgam<-qgam(
  Prevalence_females ~
    Year +
    Country+
    s(HDI, bs = 'cr') +
    Affordability,
  data=data_gam,
  qu=0.5
  #  try to remove the outliers with respect to HDI- but that does not change much
  # data= data_gam[!data_gam$Country %in% c("Mexico", "Chile", "Colombia","Costa Rica"), ]
)
```

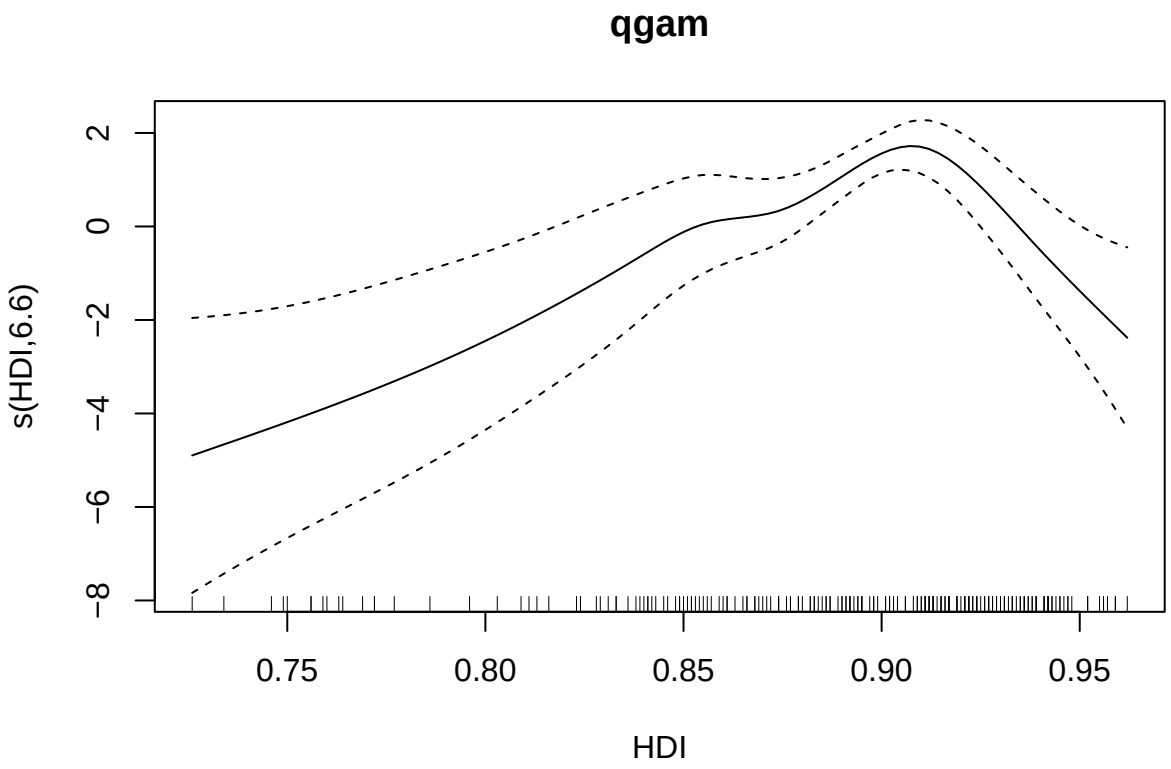
```
## Estimating learning rate. Each dot corresponds to a loss evaluation.
## qu = 0.5........done
```

```
summary(final_model_females_qgam)
```

```
##
## Family: elf
## Link function: identity
##
## Formula:
## Prevalence_females ~ Year + Country + s(HDI, bs = "cr") + Affordability
##
## Parametric coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)            597.46108   83.77119   7.132 9.89e-13 ***
## Year                    -0.28908    0.04196  -6.890 5.57e-12 ***
## CountryAustria          13.40125    0.90005  14.889  < 2e-16 ***
## CountryBelgium           6.19564    0.56225  11.019  < 2e-16 ***
## CountryCanada           -1.11340    0.50804  -2.192 0.028411 *
## CountryChile            16.75882    1.32076  12.689  < 2e-16 ***
## CountryColombia         -4.50528    1.83489  -2.455 0.014075 *
## CountryCosta Rica       -5.45045    1.48370  -3.674 0.000239 ***
## CountryCzech Republic   11.94779    0.84078  14.210  < 2e-16 ***
## CountryDenmark           6.26451    0.56327  11.122  < 2e-16 ***
## CountryEstonia           8.48249    0.87073   9.742  < 2e-16 ***
## CountryFinland           3.34654    0.54057   6.191 5.99e-10 ***
## CountryFrance           16.30567    0.80872  20.162  < 2e-16 ***
## CountryGermany           8.57802    0.46858  18.307  < 2e-16 ***
```

```
## CountryGreece           19.39888    0.94467  20.535  < 2e-16 ***
## CountryHungary          15.88651    1.20083  13.230  < 2e-16 ***
## CountryIceland           1.44983    0.51016   2.842 0.004484 **
## CountryIreland           7.66528    0.48941  15.662  < 2e-16 ***
## CountryIsrael            0.23907    0.56531   0.423 0.672366
## CountryItaly             4.77631    0.87796   5.440 5.32e-08 ***
## CountryKorea            -9.76621    0.78841 -12.387  < 2e-16 ***
## CountryLatvia           10.04095    1.13004   8.885  < 2e-16 ***
## CountryLithuania         5.01256    1.00861   4.970 6.70e-07 ***
## CountryLuxembourg        5.31546    0.74189   7.165 7.80e-13 ***
## CountryMexico           -2.12469    1.62953  -1.304 0.192279
## CountryNetherlands       7.89093    0.48052  16.422  < 2e-16 ***
## CountryNew Zealand       1.90807    0.48319   3.949 7.85e-05 ***
## CountryNorway            7.40821    0.80778   9.171  < 2e-16 ***
## CountryPoland            9.21121    1.00258   9.187  < 2e-16 ***
## CountryPortugal          5.31661    1.20365   4.417 1.00e-05 ***
## CountrySlovak Republic  10.48955    1.27144   8.250  < 2e-16 ***
## CountrySlovenia          4.94801    0.66188   7.476 7.68e-14 ***
## CountrySpain            12.43054    0.79859  15.566  < 2e-16 ***
## CountrySweden            4.05615    0.55871   7.260 3.88e-13 ***
## CountrySwitzerland      10.39410    0.78743  13.200  < 2e-16 ***
## CountryTürkiye           6.54268    1.40188   4.667 3.06e-06 ***
## CountryUnited Kingdom    3.05201    0.58536   5.214 1.85e-07 ***
## CountryUnited States     2.21417    0.60864   3.638 0.000275 ***
## Affordability           -0.53278    0.26575  -2.005 0.044983 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##          edf Ref.df Chi.sq p-value
## s(HDI) 6.601  7.747  127.5  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.977   Deviance explained = 92.5%
## -REML = 301.22  Scale est. = 1         n = 222
```

```
plot(final_model_females_qgam,main="qgam")
```




```
#plot(final_model_females_qgam, scale = FALSE, pages = 1)
```

We try to predict the value for Turkey

```
prevf<-data_gam[data_gam$Country=="Türkiye",3][1]
prevf
```

```
## [1] 16.7
```

```
target<- prevf-0.3*prevf
target
```

```
## [1] 11.69
```

```
new_obs<-data.frame(Country="Türkiye",Year=2025,HDI=0.843,Affordability=3.1)
predict(final_model_females_qgam,new_obs,se=TRUE)
```

```
## $fit
##        1
## 16.52086
##
## $se.fit
##         1
## 0.4963723
```

For a lack of time, we could not try conformalized prediction intervals for quantile regression, but will try them in the future.