

# STA442 Assignment 3

Yingfei Zha

15/11/2020

## 1. Atmospheric Carbon Dioxide Concentrations Report

The data come from the website of *The Scripps CO<sub>2</sub> Program* including over 2000 atmospheric Carbon Dioxide concentrations measurements that are observed in Hawaii from August 1960 to June 2020. The variable of interest is the concentration of Carbon Dioxide in parts per million, from which we explore if it was impacted by the events such as the fall of the Berlin wall in November 1989 and the global lockdown during the COVID-19 pandemic starting in February 2020.

A Gamma model with a semi-parametric time trend is chosen to fit the data. The integrated nested Laplace approximation (INLA) is used to determine Bayesian inference. The concentrations of Carbon Dioxide  $Y_i$  follows  $\text{Gamma}(\theta, \lambda_i \theta)$  since the data are positive and continuous. The coefficient of variation of Gamma distribution  $1/\sqrt{\theta}$  has an Exponential prior with a median at 0.1, which also means the standard deviation of  $Y_i$  is 10% of the expected value of  $Y_i$ . The log of the expected CO<sub>2</sub> concentrations  $\log(\lambda_i)$  is estimated by  $X_i \beta + U(t_i; \sigma)$ , where  $X_i$  are four sinusoidal basis functions to capture seasonal effects while  $U(t_i; \sigma)$  captures the smoothed trend without seasonality. To be more specific,  $X_{i1} = \cos(\frac{2\pi t_i}{365.25})$ ,  $X_{i2} = \sin(\frac{2\pi t_i}{365.25})$ ,  $X_{i3} = \cos(\frac{2\pi t_i}{182.625})$  and  $X_{i4} = \sin(\frac{2\pi t_i}{182.625})$  are used as linear covariates, where  $t_i$  is the time of observation  $i$ . The prior of the fixed effects is set to the default prior, which is  $\beta$  follows a Normal prior with mean equals to 0 and standard deviation equals to 1000. The random effect  $U(t_i; \sigma)$  follows a second-order random walk  $\text{RW2}(\sigma^2)$ , the prior of  $\sigma$  is an Exponential prior with a median at 0.1, which means that the slope of log rate changes by 0.1 from day to day.

The statistical results are shown in **Table 1**. The plots of overall fitted values, smoothed trend and sample derivatives of CO<sub>2</sub> from 1960 to 2021 are shown in **Figure 1**. The seasonality of data is clearly illustrated in the plot of fitted values. The plot of smoothed time trend shows an overall increasing trend of CO<sub>2</sub>. The derivatives of the time trend from 100 posterior samples are greater than 0, which indicates that atmospheric CO<sub>2</sub> concentrations were increasing all the time. Moreover, there is a significant trough in the derivatives plot in the early 90s. It could be the impact of the fall of the Berlin wall or the fall in industrial production in the Soviet Union and Eastern Europe. More detailed plots about the changes of CO<sub>2</sub> concentrations before and after the fall of the Berlin wall (1989) are shown in **Figure 2**. In the plot of fitted values, it is not obvious that the atmospheric CO<sub>2</sub> concentrations changed significantly right after the fall of the Berlin wall. The smoothed time trend of CO<sub>2</sub> concentrations was briefly flattened after the fall of the Berlin wall but it was still going up. Although the trend was flatter, the rate of change of CO<sub>2</sub> concentrations did not decrease. Instead, it went up briefly right after the fall of the Berlin wall. Then it was followed by a rapid drop in the slope of CO<sub>2</sub> trend starting from June 1991 to January 1993, it was most likely influenced by the fall in industrial production in the Soviet Union as well as Eastern Europe. On the other hand, the fall of the Berlin wall has not slowed down the increase of CO<sub>2</sub> concentrations in atmosphere right after it happened. For the changes of CO<sub>2</sub> concentrations before and after the COVID-19 lockdown starting in February 2020, more detailed plots are shown in **Figure 3**. In the plots of fitted values and smoothed time trend, they both shows that the concentration of CO<sub>2</sub> was still increasing after the global lockdown. The slope of CO<sub>2</sub> concentrations trend is not clearly increasing or decreasing after February 2020. Since there are fewer data after February 2020, the variations among the posterior samples are big. The plot of sample derivatives does

not show any evidence of the CO<sub>2</sub> concentrations growing more slowly. Maybe it is too early to conclude the impact of the COVID-19 lockdown to the CO<sub>2</sub> concentrations in air.

In conclusion, the data from the website of *The Scripps CO<sub>2</sub> Program* show that the atmospheric CO<sub>2</sub> concentrations were neither decreased by the fall of the Berlin wall in November 1989 not and the global lockdown during the COVID-19 pandemic starting in February 2020. The CO<sub>2</sub> concentrations have a constantly increasing time trend. There is no evidence to say that the COVID-19 lockdown has slowed down the increase of CO<sub>2</sub> or even speed up the increase. However, the increase of CO<sub>2</sub> was briefly accelerated right after the fall of the Berlin wall, but it was decelerated during the fall in industrial production in the Soviet Union and Eastern Europe.

Table 1: Estimated rate ratio and exponentiated SD for CO<sub>2</sub> data

	0.5quant	0.025quant	0.975quant
(Intercept)	370.033	369.708	370.358
sin12	1.008	1.008	1.008
cos12	0.998	0.997	0.998
sin6	0.998	0.998	0.998
cos6	1.002	1.002	1.002
sd for gamma	1.000	1.000	1.000
sd for timeInla	1.002	1.002	1.002

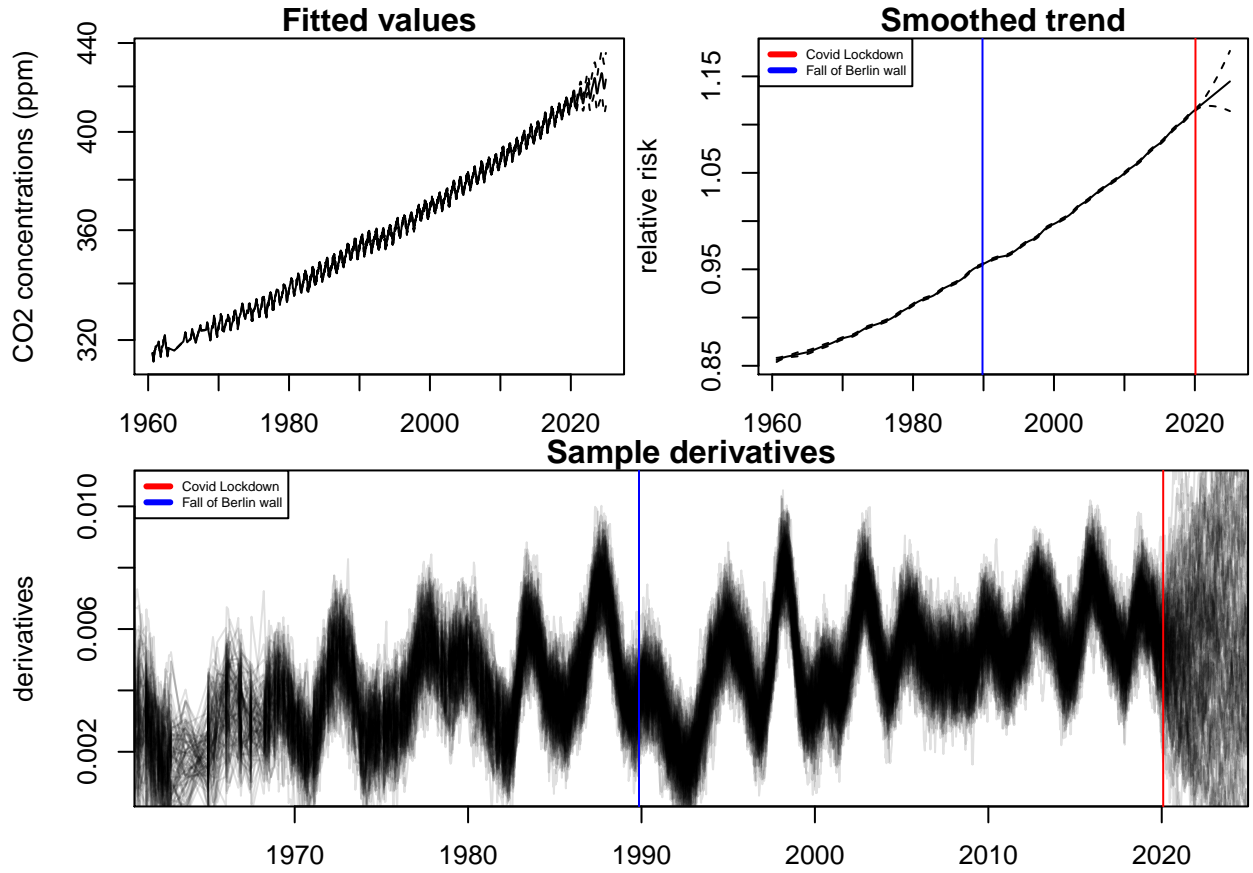


Figure 1: Plots of fitted values, smoothed trend and sample derivatives of CO<sub>2</sub> from 1960 to 2021

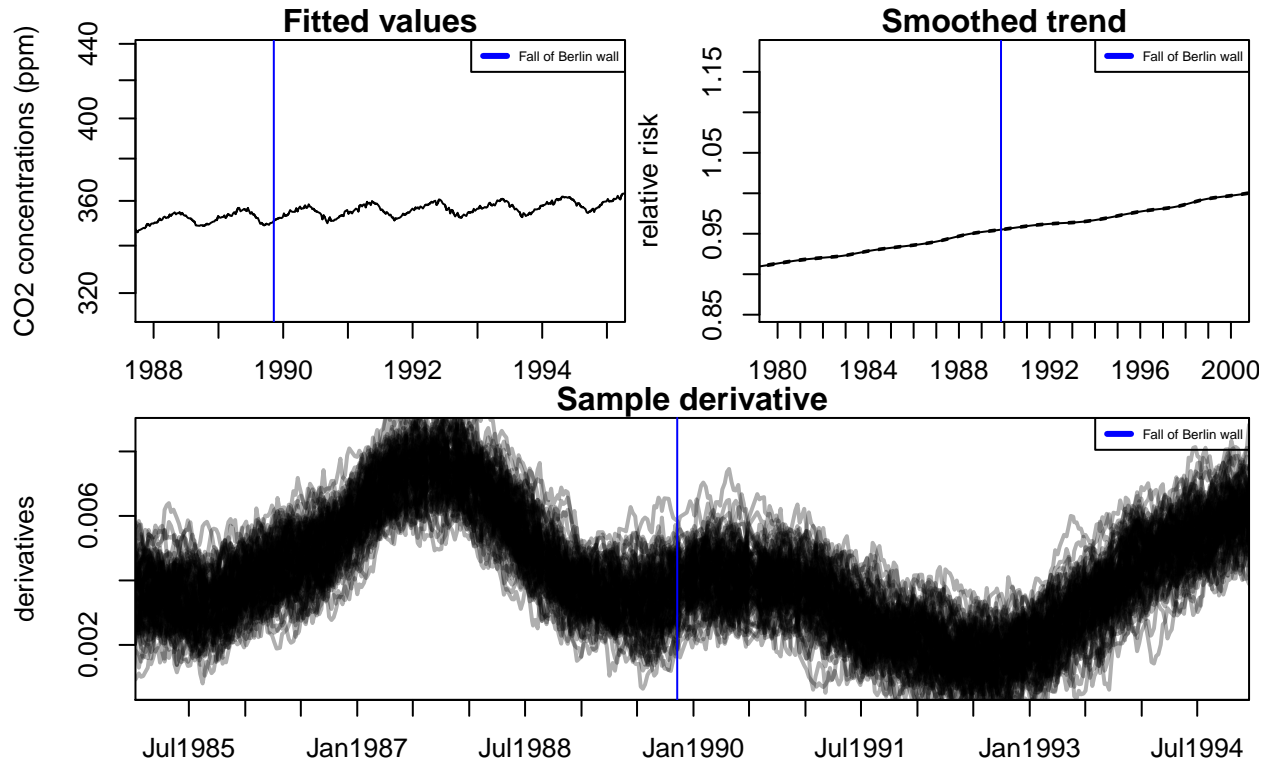


Figure 2: Plots of fitted values, smoothed trend and posterior sample derivatives of CO2 before and after the fall of the Berlin wall

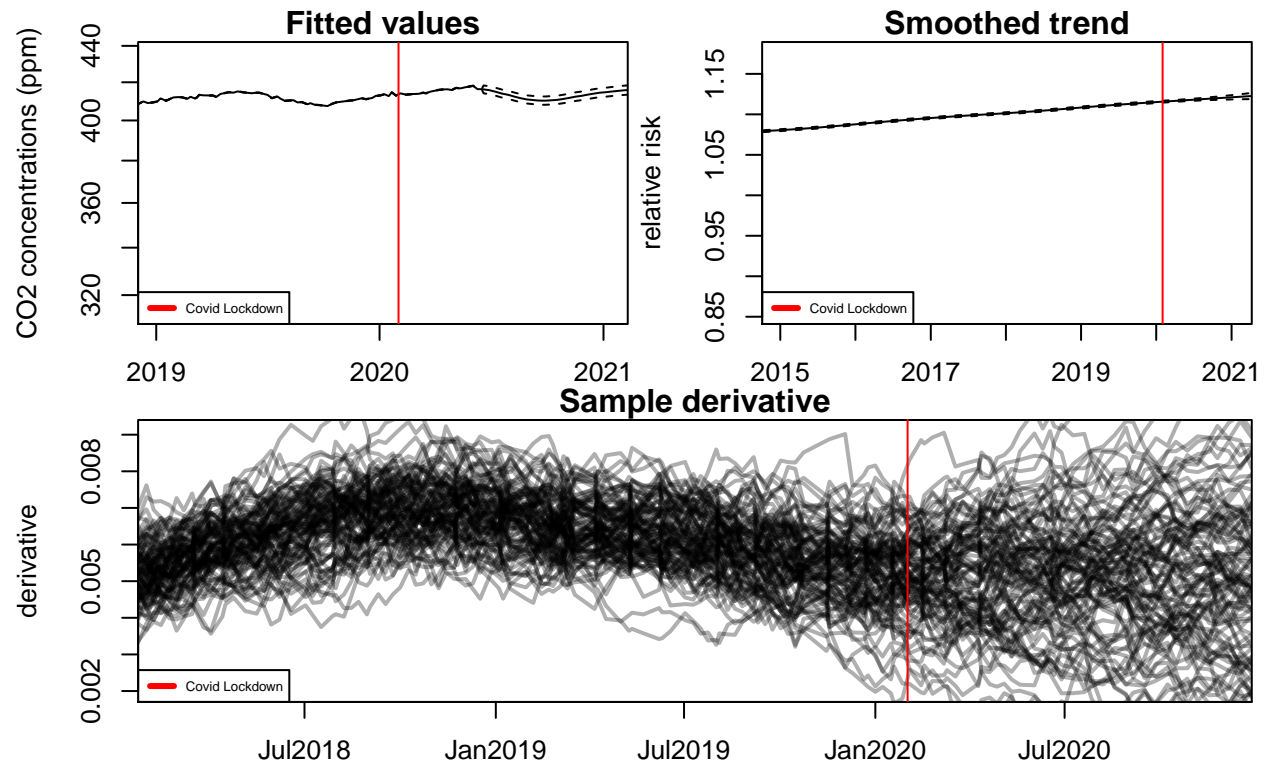


Figure 3: Plots of fitted values, smoothed trend and posterior sample derivatives of CO2 before and after the global lockdown during the COVID-19 pandemic

## 2. Report of Mortality Counts in Quebec during the COVID-19 Pandemic

### Summary

In this study, we investigate the impact of the COVID-19 epidemic on the elderly (70 years old and over) and young people (under 50 years old) in the spring and fall of 2020 respectively according to the weekly deaths data from all causes published by the Institut de la statistique du Québec (ISQ) from January 1st, 2010 to October 21st, 2020. The results suggest that, from March to the end of May, the first wave of the COVID-19 epidemic caused much more deaths of people who are 70 years old and over whereas people under 50 years old did not have a lot more deaths than what would be expected. On the other hand, the elderly had fewer deaths after September and so did the under 50's. There is not so much difference between the observed deaths and the expected number of deaths for the elderly. The under 50's even have fewer observed deaths than what would be expected. Therefore, there is no evidence suggesting that young people are the cause of the second wave of the COVID-19 epidemic.

### Introduction

In this study, the data come from the Institut de la statistique du Québec (ISQ) including the weekly numbers of deaths from all causes from January 1st, 2010 to October 21st, 2020. The mortality counts are presented by three age groups, which are 0-49 years old, 50-69 years old, 70 years old and over. The purpose of this study is to investigate how the elderly (70 years old and over) and young people (under 50 years old) are affected by the COVID-19 epidemic in the spring and fall of 2020. There are hypotheses suggesting that the first wave of the COVID-19 epidemic caused much more deaths of the elderly while not so many deaths of young people in March, April and May. Another hypothesis is that the second wave is caused by young people since the mortality counts of the under 50's has been increasing starting from September while the over 70's have no more deaths than would be expected. The analysis is based on the comparison between the observed death counts and the normally expected death counts of the above two age groups respectively.

### Methods

To investigate the above hypotheses, an overdispersed Poisson model with a semi-parametric time trend is chosen to fit the data. The integrated nested Laplace approximation (INLA) is used to determine Bayesian inference. The two models for 0-49 years old, 70 years old and over age groups are the same, except they fit different data of each age group. Besides, the data fitted are historical data before March 2020, so the fitted values since March 2020 are forecast values or the normally expected death counts. The relationship between the response (weekly death count), seasonal effects and random effect is given by:

$$Y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + U(t_i) + V_i$$

where

$$\begin{aligned}\lambda_i &= E(Y_i) \\ X_{i1} &= \sin\left(\frac{2\pi t_i}{365.25}\right) \\ X_{i2} &= \sin\left(\frac{2\pi t_i}{182.625}\right) \\ X_{i3} &= \cos\left(\frac{2\pi t_i}{365.25}\right) \\ X_{i4} &= \cos\left(\frac{2\pi t_i}{182.625}\right)\end{aligned}$$

$$U(t_i; \sigma_U) \sim \text{RW2}(\sigma_U^2)$$

$$V_i \sim \text{N}(0, \sigma_V^2)$$

The covariates  $X_i$  are four sinusoidal basis functions to capture seasonal effects and  $t_i$  is the time (in weeks) of observation  $i$ . The prior of seasonal effects  $X_i\beta$  is set to the default prior, which is  $\beta$  follows a Normal prior with mean equals to 0 and standard deviation equals to 1000. The random effect  $U(t_i; \sigma_U)$  captures the smoothed time trend without seasonality. It follows a second-order random walk  $\text{RW2}(\sigma_U^2)$ . The prior of  $\sigma_U$  is an Exponential prior with a median at 0.01, which means that the slope of log rate changes by 0.01 from week to week. The independent variance  $V_i$  covers over-dispersion of the model and it follows a Normal distribution with a zero mean and variance  $\sigma_V^2$ . The prior of  $\sigma_V$  is an Exponential prior with a median at  $\log(1.2)$ , which means that there could be 20% change in the number of deaths from week to week.

## Results

### 0-49 years old

In **Figure 4** plot 1, the seasonally adjusted trend is decreasing from 2010 to 2020 and it is expected to keep decreasing after 2020. To take a closer look at more recent data, in **Figure 4** plot 2, 30 posterior samples between June 2019 and October 2020 are chosen and their means and 95% envelop are plotted. There is not clearly a difference between the observed death counts and the normally expected death counts. Instead, most observed data of weekly deaths are included in the samples. To answer the hypothesis about the post-covid mortality counts of the under 50's, data from March to October of 2020 are shown in **Figure 5**. The excess deaths in **Figure 5** plot 2 are obtained by subtracting the fitted values of each posterior sample from the observed deaths shown in **Figure 5** plot 1. In the spring from March to May, the excess deaths fluctuate approximately between 0-40 deaths for each week. The total excess deaths in spring has a median of 92 deaths as shown in **Table 3**. In the fall starting from September, the excess deaths are approximately in between -20-30 deaths for each week and the number even go below 0 at the end of October. This indicates that the observed deaths of young people are actually fewer than the number of death normally expected. The total excess deaths in the most recent week has a median of -4 deaths as shown in **Table 4**, so the observed deaths of young people are 4 deaths fewer than the number of death normally expected in the week ends on October 24, 2020. Therefore, the hypothesis is not true about that young people have more deaths since September so that they are the cause of the second wave. In addition to this, there was an increase in the number of deaths of young people in the first wave. This increase was around 100 total deaths in between March and May.

### 70 years old and over

On the other hand, the number of deaths of the over 70's shows different trend than the under 50's. In **Figure 6** plot 1, the seasonally adjusted trend is increasing from 2010 to 2020 meaning the number of deaths is increasing generally over the years. After 2020, the number of deaths is expected to decrease. To take a closer look at more recent data, in **Figure 6** plot 2, 30 posterior samples between June 2019 and October 2020 are chosen. There is a big increase of the observed death counts comparing to the normally expected death counts in the spring. To answer the hypothesis about the post-covid mortality counts of the over 70's, data from March to October of 2020 are shown in **Figure 7**. The excess deaths in **Figure 7** plot 2 are obtained by subtracting the fitted values of each posterior sample from the observed deaths shown in **Figure 7** plot 1. In the spring from March to May, the excess deaths increase from around -100 deaths to 700 deaths for each week. The total excess deaths in spring has a median of 4616 deaths as shown in **Table 6**. Moreover, in the fall starting from September, the excess deaths are approximately in between 0-200 deaths for each week but the number is still going up at the end of October. The total excess deaths in the most recent week has a median of 144 more deaths as shown in **Table 7**, so the number of excess deaths of the elderly is much fewer than the number in spring. Therefore, the hypothesis is true about that the over 70's were greatly affected by COVID-19 in the spring of 2020. The situation got better for the elderly in the

second wave, the number of excess deaths is around 144 deaths which is not so much difference between the observed deaths and the expected number of deaths.

Table 2: Estimated rate ratio and exponentiated SD in the model of the people under 50 yrs old

	0.5quant	0.025quant	0.975quant
(Intercept)	62.086	61.335	62.839
sin12	0.989	0.973	1.005
sin6	1.013	0.996	1.029
cos12	0.989	0.973	1.005
cos6	1.038	1.021	1.054
sd for timeIid	1.044	1.024	1.066
sd for timeForInla	1.013	1.003	1.038

Table 3: Estimated rate ratio and exponentiated SD in the model of the people over 70 yrs old

	0.5quant	0.025quant	0.975quant
(Intercept)	884.655	877.941	891.151
sin12	1.066	1.058	1.074
sin6	1.011	1.005	1.018
cos12	1.124	1.116	1.133
cos6	1.012	1.005	1.018
sd for timeIid	1.041	1.037	1.046
sd for timeForInla	1.174	1.113	1.271

Table 4: Quantiles of total excess deaths from the 30 posterior samples of people under 50 yrs old from March to May

Number of excess deaths	
0%	1
25%	74
50%	112
75%	136
100%	177

Table 5: Quantiles of total excess deaths from the 30 posterior samples of people under 50 yrs old in the most recent week

Number of excess deaths	
0%	-19
25%	-11
50%	-4
75%	4
100%	9

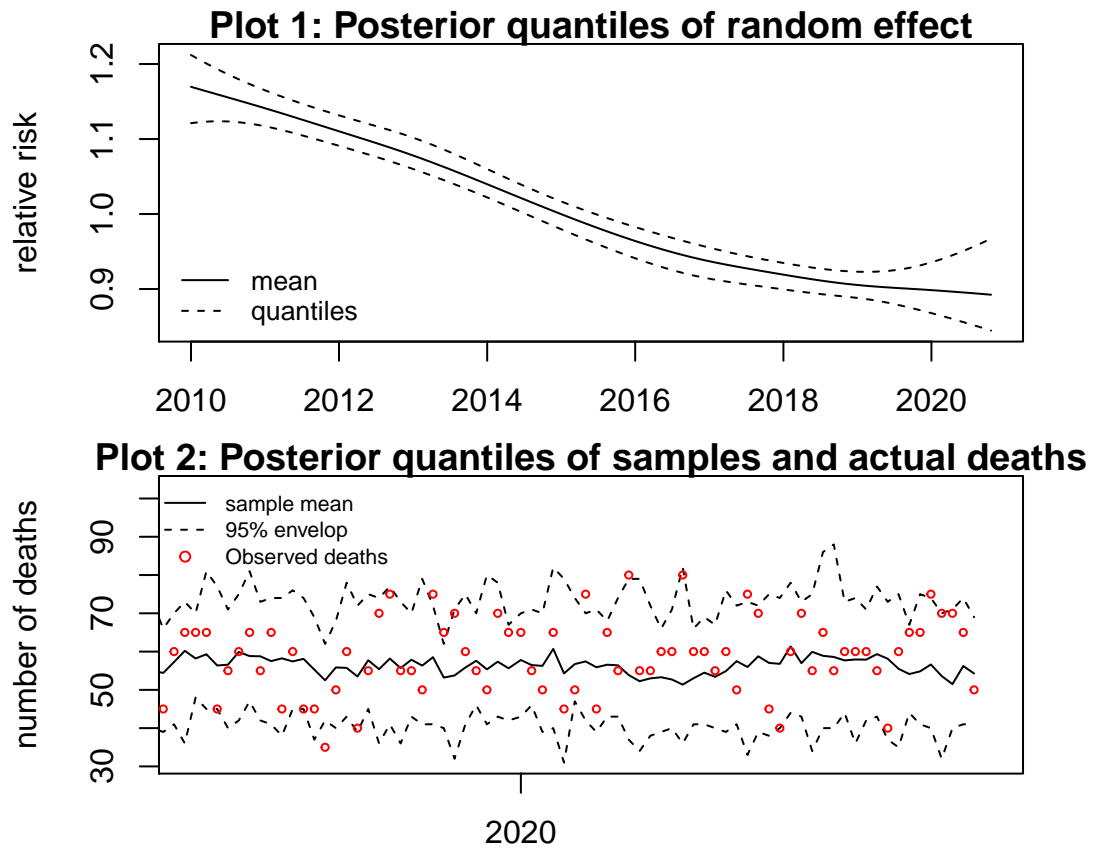


Figure 4: Plots of seasonally adjusted trend and posterior quantiles of 30 posterior samples with the observed deaths of young people.

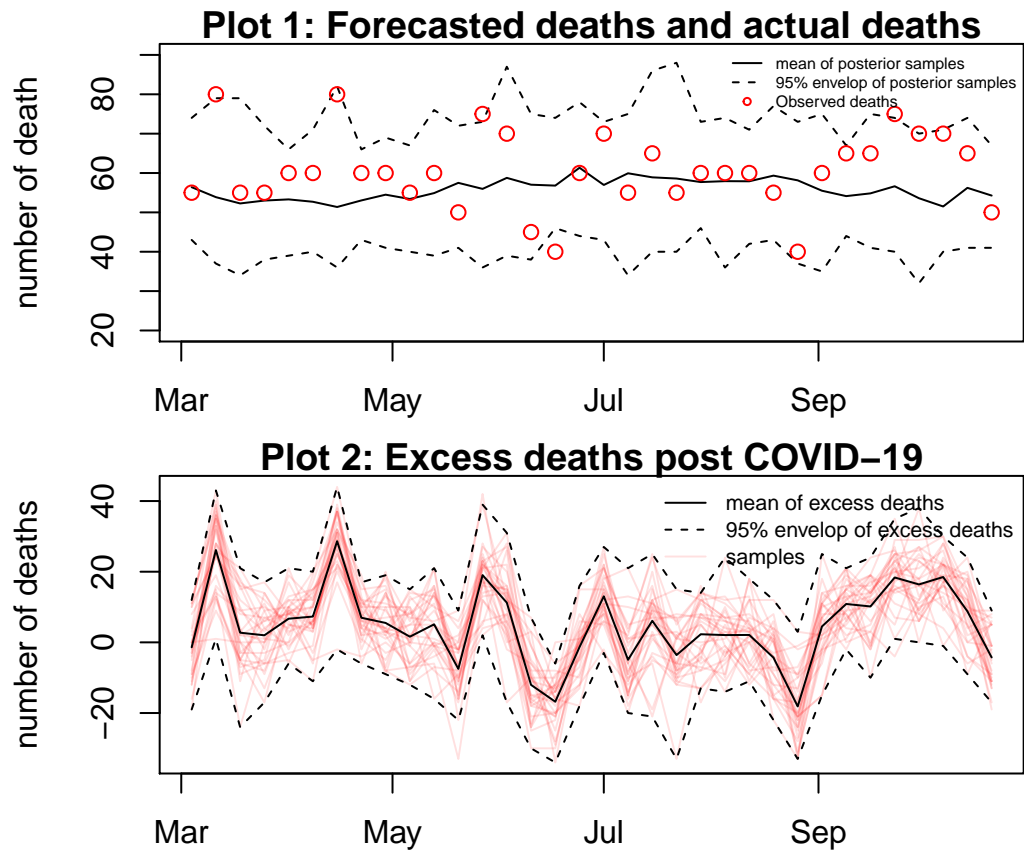


Figure 5: Plots of posterior quantiles of samples with the observed deaths and excess deaths of young people since March 2020.



Table 6: Quantiles of total excess deaths from the 30 posterior samples of people over 70 yrs old from March to May

Number of excess deaths	
0%	3724
25%	4373
50%	4610
75%	4777
100%	5373

Table 7: Quantiles of total excess deaths from the 30 posterior samples of people over 70 yrs old in the most recent week

Number of excess deaths	
0%	-59
25%	108
50%	157
75%	216
100%	345

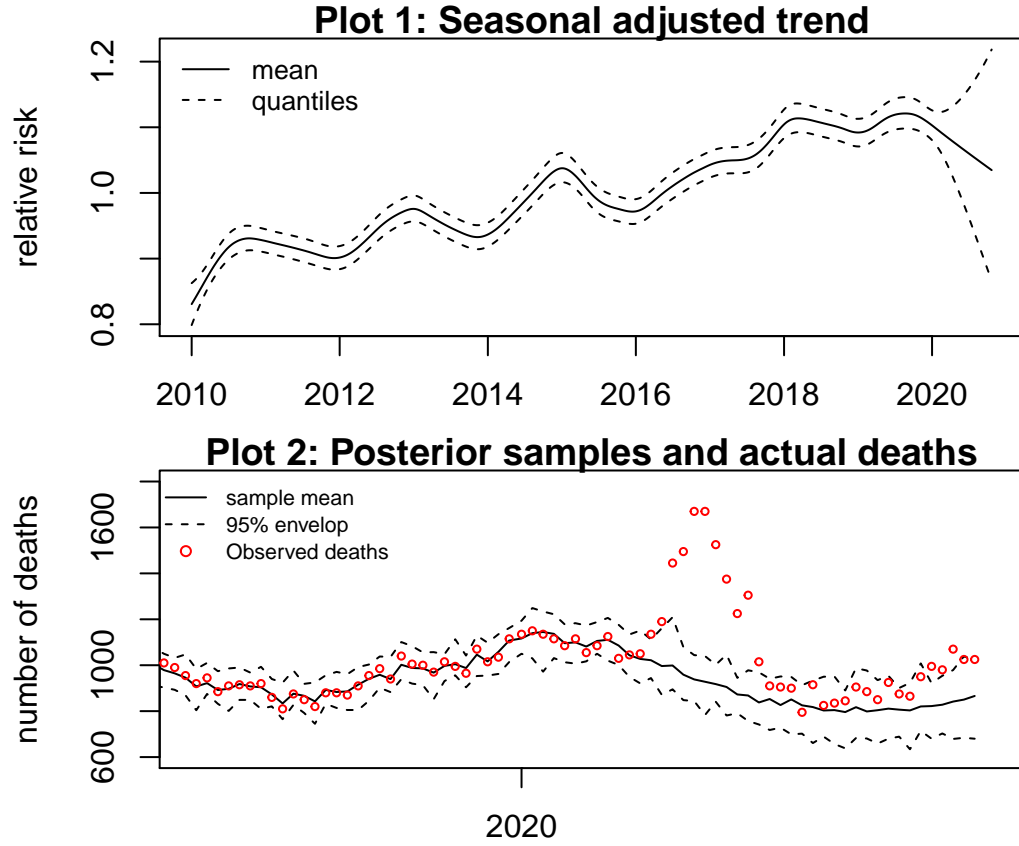


Figure 6: Plots of seasonally adjusted trend and posterior quantiles of 30 posterior samples with the observed deaths of the elderly.

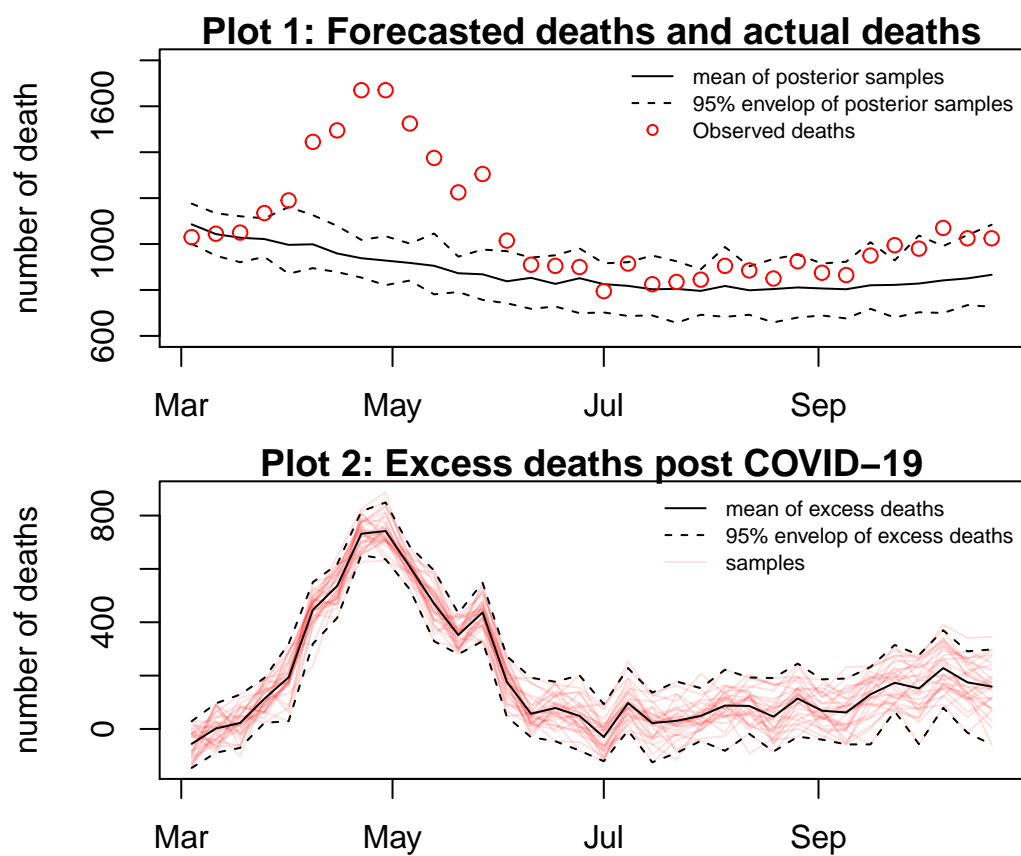


Figure 7: Plots of posterior quantiles of the samples with the observed deaths and excess deaths of the elderly since March 2020.

## Appendix

```
knitr::opts_chunk$set(echo=FALSE, warning = FALSE, message = FALSE)
cUrl = paste0("http://scrippsco2.ucsd.edu/assets/data/atmospheric/", "stations/flask_co2/daily/daily_fl")
cFile = basename(cUrl)
if (!file.exists(cFile)) download.file(cUrl, cFile)
co2s = read.table(cFile, header = FALSE, sep = ",", skip = 69, stringsAsFactors = FALSE,
                  col.names = c("day", "time", "junk1", "junk2", "Nflasks", "quality", "co2"))
co2s$date = strptime(paste(co2s$day, co2s$time), format = "%Y-%m-%d %H:%M", tz = "UTC")
# remove low-quality measurements
co2s = co2s[co2s$quality == 0, ]

plot(co2s$date, co2s$co2, log = "y", cex = 0.3, col = "#00000040", xlab = "time", ylab = "ppm")
plot(co2s[co2s$date > ISOdate(2015, 3, 1, tz = "UTC"), c("date", "co2")], log = "y", type = "o", xlab =
# forecast
co2s$day = as.Date(co2s$date)
toAdd = data.frame(day = seq(max(co2s$day) + 3, as.Date("2025/1/1"), by = "10 days"), co2 = NA)
co2ext = rbind(co2s[, colnames(toAdd)], toAdd)
timeOrigin = as.Date("2000/1/1")
co2ext$timeInla = round(as.numeric(co2ext$day - timeOrigin)/365.25, 2)
co2ext$cos12 = cos(2 * pi * co2ext$timeInla)
co2ext$sin12 = sin(2 * pi * co2ext$timeInla)
co2ext$cos6 = cos(2 * 2 * pi * co2ext$timeInla)
co2ext$sin6 = sin(2 * 2 * pi * co2ext$timeInla)
library('INLA', verbose=FALSE)
# disable some error checking in INLA
mm = get("inla.models", INLA:::inla.get.inlaEnv())
if(class(mm) == 'function') mm = mm()
mm$latent$rw2$min.diff = NULL
assign("inla.models", mm, INLA:::inla.get.inlaEnv())

co2res = inla(co2 ~ sin12 + cos12 + sin6 + cos6 + f(timeInla, model = 'rw2',
prior='pc.prec', param = c(0.1, 0.5)), data = co2ext, family='gamma', control.family = list(hyper=list(
# add this line if your computer has trouble
# control.inla = list(strategy='gaussian'),
control.predictor = list(compute=TRUE, link=1), control.compute = list(config=TRUE), verbose=FALSE)
qCols = c('0.5quant', '0.025quant', '0.975quant')
knitr::kable(rbind(exp(co2res$summary.fixed[, qCols]), exp(Pmisc::priorPost(co2res)$summary[,
qCols])), digits = 3, caption="Estimated rate ratio and exponentiated SD for CO2 data")
sampleList = INLA:::inla.posterior.sample(100, co2res, selection = list(timeInla = 0))
names(sampleList[[1]])
sampleList[[1]]$hyperpar
length(sampleList[[1]]$latent) #number of time points
sampleMean = do.call(cbind, Biobase::subListExtract(sampleList, "latent"))
sampleDeriv = apply(sampleMean, 2, diff)/diff(co2res$summary.random$timeInla$ID) #[s(t) - s(t-t)]/time
Stime = timeOrigin + round(365.25 * co2res$summary.random$timeInla$ID)
layout(matrix(c(1,2,3,3), 2, 2, byrow = TRUE))
par(mai = c(0.3, 0.65, 0.2, 0.05))
# Prediction
matplot(co2ext$day, co2res$summary.fitted.values[, qCols], type = "l", col = "black", lty = c(1, 2, 2),
# Random effect
matplot(Stime, exp(co2res$summary.random$timeInla[, qCols]),
type = "l", col = "black", lty = c(1, 2, 2), xlab = "time", ylab = "relative risk", main='Smoothed trend
```

```

abline(v = as.Date("1989-11-09"), col = "blue")
abline(v = as.Date("2020-02-01"), col = "red")
legend("topleft", col=c("red","blue"),lty=1, lwd=3, legend=c("Covid Lockdown", "Fall of Berlin wall"),cex = 0.5)
#Posterior samples random effect. 30 lines they are the same at where there are data but behave differently
#matplot(Stime, exp(sampleMean), type='l')
#Derivative
Stime = timeOrigin + round(365.25 * co2res$summary.random$timeInla$ID)
matplot(Stime[-1], sampleDeriv, type = "l", lty = 1,
xaxs = "i", col = "#00000020", xlab = "time", ylab = "derivatives",
ylim = quantile(sampleDeriv, c(0.01, 0.995)),main='Sample derivatives')
abline(v = as.Date("1989-11-09"), col = "blue")
abline(v = as.Date("2020-02-01"), col = "red")
legend("topleft", col=c("red","blue"),lty=1, lwd=3, legend=c("Covid Lockdown", "Fall of Berlin wall"),cex = 0.5)
#Zoomed in fitted value ,widths = rep.int(lcm(5), ncol(matrix(c(1,2,3,3), 2, 2, byrow = TRUE))),
layout(matrix(c(1,2,3,3), 2, 2, byrow = TRUE), heights = rep.int(lcm(5), ncol(matrix(c(1,2,3,3), 2, 2, byrow = TRUE))),
par(mai = c(0.3, 0.65, 0.2, 0.05))
matplot(co2ext$day, co2res$summary.fitted.values[,
qCols], type = "l", col = "black", lty = c(1, 2,
2), log = "y", xlab = "time", ylab = "CO2 concentrations (ppm)",
xlim=c(as.Date("1988-01-01"), as.Date("1995-01-01")), xaxt='n', main='Fitted values')
Saxis = seq(as.Date("1988-01-01"), as.Date("1995-01-01"), by = '1 year')
axis(1,at=as.numeric(Saxis),labels=format(Saxis, "%Y"))
abline(v = as.Date("1989-11-09"), col = "blue")
legend("topright", col=c("blue"),lty=1, lwd=3, legend=c("Fall of Berlin wall"),cex = 0.5)
#Zoomed in smoothed trend
matplot(Stime, exp(co2res$summary.random$timeInla[, qCols]),
type = "l", col = "black", lty = c(1, 2, 2), xlab = "time", ylab = "relative risk", xlim=c(as.Date("1980-01-01"), as.Date("2000-01-01")),
Saxis = seq(as.Date("1980-01-01"), as.Date("2000-01-01"), by = '1 year')
axis(1,at=as.numeric(Saxis),labels=format(Saxis, "%Y"))
abline(v = as.Date("1989-11-09"), col = "blue")
legend("topright", col=c("blue"),lty=1, lwd=3, legend=c("Fall of Berlin wall"),cex = 0.5)
#Zoomed in derivative
forX = as.Date(c("1985/1/1", "1995/1/1"))
forX = seq(forX[1], forX[2], by = "6 months")
toPlot = which(Stime > min(forX) & Stime < max(forX))
matplot(Stime[toPlot], sampleDeriv[toPlot, ], type = "l",
lty = 1, lwd = 2, xaxs = "i", col = "#00000050",
xlab = "time", ylab = "derivatives", xaxt = "n", ylim = quantile(sampleDeriv[toPlot,
], c(0.01, 0.995)), main='Sample derivative')
axis(1, as.numeric(forX), format(forX, "%b%Y"))
abline(v = as.Date("1989-11-09"), col = "blue")
legend("topright", col="blue",lty=1, lwd=3, legend="Fall of Berlin wall",cex = 0.5)

layout(matrix(c(1,2,3,3), 2, 2, byrow = TRUE), heights = rep.int(lcm(5), ncol(matrix(c(1,2,3,3), 2, 2, byrow = TRUE))),
par(mai = c(0.3, 0.65, 0.2, 0.05))
matplot(co2ext$day, co2res$summary.fitted.values[,
qCols], type = "l", col = "black", lty = c(1, 2,
2), log = "y", xlab = "time", ylab = "CO2 concentrations (ppm)",
xlim=c(as.Date("2019-01-01"), as.Date("2021-01-11")),xaxt='n', main='Fitted values')
Saxis = seq(as.Date("2019-01-01"), as.Date("2021-01-11"), by = '1 year')
axis(1,at=as.numeric(Saxis),labels=format(Saxis, "%Y"))
abline(v = as.Date("2020-02-01"), col = "red")
legend("bottomleft", col=c("red"),lty=1, lwd=3, legend=c("Covid Lockdown"),cex = 0.5)

```

```

#Zoomed in smoothed trend
matplot(Stime, exp(co2res$summary.random$timeInla[, qCols]),
type = "l", col = "black", lty = c(1, 2, 2), xlab = "time", ylab = "relative risk", xlim=c(as.Date("2015-01-01"), as.Date("2021-01-11")),
Saxis = seq(as.Date("2015-01-01"), as.Date("2021-01-11"), by = '1 year')
axis(1,at=as.numeric(Saxis),labels=format(Saxis, "%Y"))
abline(v = as.Date("2020-02-01"), col = "red")
legend("bottomleft", col=c("red"),lty=1, lwd=3, legend=c("Covid Lockdown"),cex = 0.5)

forX = as.Date(c("2018/1/1", "2021/1/1"))
forX = seq(forX[1], forX[2], by = "6 months")
toPlot = which(Stime > min(forX) & Stime < max(forX))
matplot(Stime[toPlot], sampleDeriv[toPlot, ], type = "l",
lty = 1, lwd = 2, xaxs = "i", col = "#00000050",
xlab = "time", ylab = "derivative", xaxt = "n", ylim = quantile(sampleDeriv[toPlot, ], c(0.01, 0.995)), main='Sample derivative')
axis(1, as.numeric(forX), format(forX, "%b%Y"))
abline(v = as.Date("2020-02-01"), col = "red")
legend("bottomleft", col=c("red"),lty=1, lwd=3, legend=c("Covid Lockdown"),cex = 0.5)
xWide = read.table(paste0("https://www.stat.gouv.qc.ca/statistiques/", "population-demographie/deces-mortalite", "age", paste0("w", 1:53)))
xWide = xWide[grepl("^[:digit:]]+$", xWide$year), ]
x = reshape2::melt(xWide, id.vars = c("year", "age"), measure.vars = grepl("^w[:digit:]]+$", colnames(xWide)),
x$dead = as.numeric(gsub("[:space:]", "", x$value))
x$week = as.numeric(gsub("w", "", x$variable))
x$year = as.numeric(x$year)
x = x[order(x$year, x$week, x$age), ]

newYearsDay = as.Date(ISOdate(x$year, 1, 1))
x$time = newYearsDay + 7 * (x$week - 1)
x = x[!is.na(x$dead), ]
x = x[x$week < 53, ]

#Divide the data into pre and post covid, add extra dates to data so that INLA will create forecasts.
dateCutoff = as.Date("2020/3/1")
xPreCovid = x[x$time < dateCutoff, ]
xPostCovid = x[x$time >= dateCutoff, ]
toForecast = expand.grid(age = unique(x$age), time = unique(xPostCovid$time), dead = NA)
xForInla = rbind(xPreCovid[, colnames(toForecast)], toForecast)
xForInla = xForInla[order(xForInla$time, xForInla$age),]

#Create some time variables, including sines and cosines. Time in years and centred is numerically stable
xForInla$timeNumeric = as.numeric(xForInla$time)
xForInla$timeForInla = (xForInla$timeNumeric - as.numeric(as.Date("2015/1/1")))/365.25
xForInla$timeId = xForInla$timeNumeric
xForInla$sin12 = sin(2 * pi * xForInla$timeNumeric/365.25)
xForInla$sin6 = sin(2 * pi * xForInla$timeNumeric * 2/365.25)
xForInla$cos12 = cos(2 * pi * xForInla$timeNumeric/365.25)
xForInla$cos6 = cos(2 * pi * xForInla$timeNumeric * 2/365.25)
plot(x[x$age == "0-49 years old", c("time", "dead")], type = "o",
log = "y")
xWide2 = reshape2::dcast(x, week + age ~ year, value.var = "dead")
Syear = grep("[:digit:]", colnames(xWide2), value = TRUE)
Scol = RColorBrewer::brewer.pal(length(Syear), "Spectral")
matplot(xWide2[xWide2$age == "0-49 years old", Scol], type = "l",
lty = 1, col = Scol, xlab="week")

```

```

legend("topright", col = Scol, legend = Syear, bty = "n",
lty = 1, lwd = 3)
xForInlaYoung= xForInla[xForInla$age == '0-49 years old', ]
library(INLA, verbose=FALSE)
mod.young = inla(dead ~ sin12 + sin6 + cos12 + cos6 +
f(timeId, prior='pc.prec', param= c(log(1.2), 0.5)) +
f(timeForInla, model = 'rw2', prior='pc.prec', param= c(0.01, 0.5)),
data=xForInlaYoung,
control.predictor = list(compute=TRUE, link=1),
control.compute = list(config=TRUE),
# control.inla = list(fast=FALSE, strategy='laplace'),
family='poisson')
qCols = paste0(c(0.5, 0.025, 0.975), "quant")
knitr::kable(rbind(exp(mod.young$summary.fixed[, qCols]), exp(Pmisc::priorPost(mod.young)$summary[,
qCols])), digits = 3, caption="Estimated rate ratio and exponentiated SD in the model of the people under 50",
set.seed(1))
sampleList.young = INLA::inla.posterior.sample(30, mod.young, selection = list(Predictor = 0))
sampleIntensity.young = exp(do.call(cbind, Biobase::subListExtract(sampleList.young,
"latent")))
sampleDeaths.young = matrix(rpois(length(sampleIntensity.young),
sampleIntensity.young), nrow(sampleIntensity.young), ncol(sampleIntensity.young))
xPostCovidYoung = xPostCovid[xPostCovid$age == "0-49 years old",
]
xPostCovidForecastYoung = sampleDeaths.young[match(xPostCovidYoung$time,
xForInlaYoung$time), ]
excessDeathsYoung = xPostCovidYoung$dead - xPostCovidForecastYoung
plot(x[x$age == "70 years old and over", c("time", "dead")], type = "o",
log = "y")
xWide2 = reshape2::dcast(x, week + age ~ year, value.var = "dead")
Syear = grep("[[:digit:]]", colnames(xWide2), value = TRUE)
Scol = RColorBrewer::brewer.pal(length(Syear), "Spectral")
matplot(xWide2[xWide2$age == "70 years old and over", Syear], type = "l",
lty = 1, col = Scol, xlab="week")
legend("topright", col = Scol, legend = Syear, bty = "n",
lty = 1, lwd = 3, cex = 0.5)
xForInlaOld= xForInla[xForInla$age == "70 years old and over", ]
library(INLA, verbose=FALSE)
mod.old = inla(dead ~ sin12 + sin6 + cos12 + cos6 +
f(timeId, prior='pc.prec', param= c(log(1.2), 0.5)) +
f(timeForInla, model = 'rw2', prior='pc.prec', param= c(0.01, 0.5)),
data=xForInlaOld,
control.predictor = list(compute=TRUE, link=1),
control.compute = list(config=TRUE),
# control.inla = list(fast=FALSE, strategy='laplace'),
family='poisson')
qCols = paste0(c(0.5, 0.025, 0.975), "quant")
knitr::kable(rbind(exp(mod.old$summary.fixed[, qCols]), exp(Pmisc::priorPost(mod.old)$summary[,
qCols])), digits = 3, caption="Estimated rate ratio and exponentiated SD in the model of the people over 70",
set.seed(1))
sampleList.old = INLA::inla.posterior.sample(30, mod.old, selection = list(Predictor = 0))

sampleIntensity.old = exp(do.call(cbind, Biobase::subListExtract(sampleList.old,
"latent")))
sampleDeaths.old = matrix(rpois(length(sampleIntensity.old),

```



```

sampleIntensity.old), nrow(sampleIntensity.old), ncol(sampleIntensity.old))
xPostCovidOld = xPostCovid[xPostCovid$age == "70 years old and over",
]
xPostCovidForecastOld = sampleDeaths.old[match(xPostCovidOld$time,
xForInlaOld$time), ]
excessDeathsOld = xPostCovidOld$dead - xPostCovidForecastOld

par(mfrow=c(2,1))
par(mai = c(0.5, 1, 0.2, 1))
#time trend of the young all the time
matplot(xForInlaYoung$time,exp(mod.young$summary.random$timeForInla[, c("0.5quant", "0.975quant", "0.025quant")]),
legend("bottomleft", col=c("black","black"),bty="n", legend=c("mean", "quantiles"), lty=c(1,2),cex = 0.5)

#Young samples real post covid
library(GET)
cset.y = GET::create_curve_set(list(r=as.numeric(xForInlaYoung$time), obs=sampleDeaths.young))
myEnv.y = GET::central_region(cset.y, coverage=0.95)
matplot(xForInlaYoung$time, as.data.frame(myEnv.y)[,c("lo","hi","central")], type = "l", lty = c(2, 2, 1),
"2020/11/1"), main="Plot 2: Posterior quantiles of samples and actual deaths")
points(x[x$age == "0-49 years old", c("time", "dead")], col = "red",
cex = 0.5)
legend("topleft", col=c("black","black","red"), bty="n", legend=c("sample mean", "95% envelop","Observed"))

par(mfrow=c(2,1))
par(mai = c(0.5, 1, 0.2, 1))
#Young sample forecast vs real post covid
cset2.y = GET::create_curve_set(list(r=as.numeric(xPostCovidYoung$time), obs=xPostCovidForecastYoung))
myEnv2.y = GET::central_region(cset2.y, coverage=0.95)
matplot(xPostCovidYoung$time, as.data.frame(myEnv2.y)[,c("lo","hi","central")], type = "l", lty = c(2, 2, 1),
points(xPostCovidYoung[, c("time", "dead")], col = "red")
legend("topright", col=c("black","black","red"), bty="n", legend=c("mean of posterior samples", "95% envelop","Observed"))

#Young excess deaths post covid
cset3.y = GET::create_curve_set(list(r=as.numeric(xPostCovidYoung$time), obs=excessDeathsYoung))
myEnv3.y = GET::central_region(cset3.y, coverage=0.95)
matplot(xPostCovidYoung$time, excessDeathsYoung, type = "l", lty = 1, col = "#FF000020", xlab='time', ylab='excess deaths')
matlines(xPostCovidYoung$time, as.data.frame(myEnv3.y)[,c("lo","hi","central")], type = "l", lty = c(2, 2, 1),
legend("topright", col=c("black","black", "#FF000020"), bty="n", legend=c("mean of excess deaths", "95% envelop","Observed"))
#Total excess deaths march-may inclusive
excessDeathsSubYoung = excessDeathsYoung[xPostCovidYoung$time >
as.Date("2020/03/01") & xPostCovidYoung$time <
as.Date("2020/06/01"), ]
excessDeathsInPeriodYoung = apply(excessDeathsSubYoung, 2, sum)
knitr::kable(round(quantile(excessDeathsInPeriodYoung),
col.names="Number of excess deaths", caption = "Quantiles of total excess deaths from the march-may inclusive"))
#Excess deaths in most recent week
knitr::kable(round(quantile(excessDeathsYoung[nrow(excessDeathsYoung), ]),
col.names="Number of excess deaths", caption = "Quantiles of total excess deaths from the most recent week"))
#Total excess deaths march-may inclusive
excessDeathsSubOld = excessDeathsOld[xPostCovidOld$time >
as.Date("2020/03/01") & xPostCovidOld$time <
as.Date("2020/06/01"), ]
excessDeathsInPeriodOld = apply(excessDeathsSubOld, 2, sum)

```

```

knitr::kable(round(quantile(excessDeathsInPeriodOld)), col.names="Number of excess deaths", caption = "
#Excess deaths in most recent week
knitr::kable(round(quantile(excessDeathsOld[nrow(excessDeathsOld), ])), col.names="Number of excess dea
par(mfrow=c(2,1))
par(mai = c(0.5, 1, 0.2, 1))
#time trend of the elder all the time
matplot(xForInlaOld$time,exp(mod.old$summary.random$timeForInla[, c("0.5quant", "0.975quant", "0.025qua
legend("topleft", col=c("black","black"),bty="n", legend=c("mean", "quantiles"), lty=c(1,2),cex = 0.8)
#Elders samples real post covid

library(GET)
cset = GET::create_curve_set(list(r=as.numeric(xForInlaOld$time), obs=sampleDeaths.old))
myEnv = GET::central_region(cset, coverage=0.95)
matplot(xForInlaOld$time, as.data.frame(myEnv)[,c("lo","hi","central")], type = "l", lty = c(2, 2, 1), c
"2020/11/1")),ylim=c(600, 1800), main="Plot 2: Posterior samples and actual deaths")
points(x[x$age == "70 years old and over", c("time", "dead")], col = "red",
cex = 0.5)
legend("topleft", col=c("black","black","red"), bty="n", legend=c("sample mean", "95% envelop","Observed
par(mfrow=c(2,1))
par(mai = c(0.5, 1, 0.2, 1))
#Elders sample forecast vs real post covid
cset2 = GET::create_curve_set(list(r=as.numeric(xPostCovidOld$time), obs=xPostCovidForecastOld))
myEnv2 = GET::central_region(cset2, coverage=0.95)
matplot(xPostCovidOld$time, as.data.frame(myEnv2)[,c("lo","hi","central")], type = "l", lty = c(2, 2, 1)
points(xPostCovidOld[, c("time", "dead")], col = "red")
legend("topright", col=c("black","black","red"), bty="n", legend=c("mean of posterior samples", "95% en

#Elders excess deaths post covid
cset3 = GET::create_curve_set(list(r=as.numeric(xPostCovidOld$time), obs=excessDeathsOld))
myEnv3 = GET::central_region(cset3, coverage=0.95)
matplot(xPostCovidOld$time, excessDeathsOld, type = "l", lty = 1, col = "#FF000020", xlab='time', ylab=
matlines(xPostCovidOld$time, as.data.frame(myEnv3)[,c("lo","hi","central")], type = "l", lty = c(2, 2, 1)
legend("topright", col=c("black","black", "#FF000020"), bty="n", legend=c("mean of excess deaths", "95%

```