Homework 1, Smoothing

Huifeng Wu

Historical Events with Carbon Dioxide Concentrations

Introduction

The Scripps CO_2 Program measures atmospheric Carbon Dioxide concentrations globally. In this report, we aim to investigate the impact of specific events on the CO_2 samples collected at Mauna Loa Observatory in Hawaii. First, it is of great significance to evaluate the change of CO_2 data, followed by the slump of industrial production with the Fall of the Berlin Wall. The second focus of this report is to study the effect of economic recession on the CO_2 data, due to the COVID-19 pandemic.

Method

We assume that the response CO_2, say Y_i, follows the Gamma distribution$(\theta, \frac{\lambda_i}{\theta})$, for some shape parameter θ and scale parameter $\frac{\lambda_i}{\theta}$. A generalized additive model (GAM) with a log link is proposed below to interpret the data, as follows:

$$\log(\lambda_i) = X\beta + U(t_i)$$
$$[U_1 \dots U_T]^T \sim \text{RW2}\left(0, \sigma_U^2\right)$$
$$X_{i0} = 1$$
$$X_{i1} = \cos\left(2\pi t_i/365.25\right)$$
$$X_{i2} = \sin\left(2\pi t_i/365.25\right)$$
$$X_{i3} = \cos\left(2\pi t_i/182.625\right)$$
$$X_{i4} = \sin\left(2\pi t_i/182.625\right)$$

where we used a 12 month and a 6 month frequency in modelling. Note $\mathrm{E}\left(Y_i\right) = \theta\lambda_i/\theta = \lambda_i$. X_{i1}, X_{i2}, X_{i3} and X_{i4} are four sinusoidal basis functions to take account for a wide range of seasonal effects. β indicates the coefficient estimates. In addition, $U(t)$ is a second-order random walk, which considers variations between data points.

$1/\sqrt{\theta}$ represents the coefficient of variation, in which θ is a precision with some prior distribution. Bayesian inference allowed us to integrate prior knowledge with the hyperparameters, σ_u and $1/\sqrt{\theta}$, which involve different aspects of randomness by the model. It is typically believed that the random slope, which is the log rate of change in CO_2 emission, could fluctuate within a range of 0.1 from time to time, so does the coefficient of variation $1/\sqrt{\theta}$. We think the standard deviations (SD) is around 10% of the the mean. Therefore, we fixed all the aforementioned hyperparameters using penalized complexity priors for precision in R-INLA, as follow:

$$P(1/\sqrt{\theta} > 0.1) = 0.5 \Longrightarrow 1/\sqrt{\theta} \sim \exp\left(10\log(2)\right)$$
$$P(\sigma_U^2 > 0.1) = 0.5 \Longrightarrow \sigma_U^2 \sim \exp\left(10\log(2)\right)$$

In this report, we take advantage of the Gamma regression to model the continuous positive CO_2 data. Not only does the model can capture a broad range of shapes, but it also give flexible choices when it comes to prior and posterior distributions. Also, confronted with such time series data of CO_2 emission, the use of first harmonics is helpful to handle the periodicity. Thus, the proposed model can capture seasonal effects on semi-annual and annual basis, and it involves natural variation thorough its underlying random effect terms.
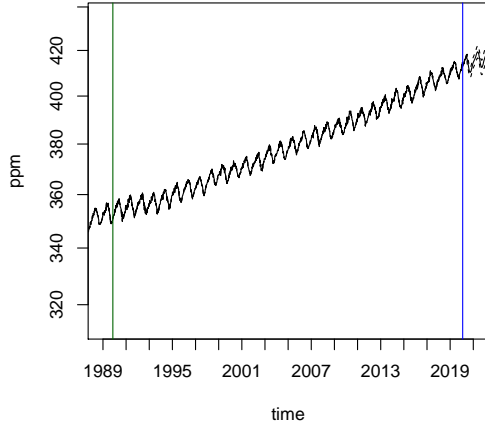
Result

In a natural scale, Table 1 delivers the results of SD for the coefficient of variation and random slope. In terms of relative rates, the change in the random slope (1.002) is generally greater than that of coefficient of variation (1.000). Both types of variation are small in the dataset. In a more piratical sense, we could expect the measurements taken at the same time would be indifferent, while movements are evident over time.

For the events of interest, we created a few figures to better assess the research hypotheses, in terms of model predication and random effect. Figure 1(a) illustrates the proposed linear model predicting the CO_2 data, and Figure 1(b) outlines the trend of random effect. In particular, vertical green and blue lines are used to timestamp the two events, the Fall of the Berlin Wall and the COVID-19 pandemic, respectively. According to Figure 1(a), we could not observe the influences of the two events on $CO2$, since the use of sinusoidal basis functions involve cyclicality naturally. Although we can see increasing trends at both time points, it is more rigorous to determine with more evaluation. Figure 1(b) shows the estimated smoothed trend of CO2. Specifically, the trend seems to appear shallower after the dramatic fall in industrial production by the Berlin Wall Fall, whereas that remains relatively steady upon COVID-19.

Moreover, it is crucial to investigate the growth rate change on CO_2 by the first order derivative of its trend. Figure 1(c) confirms our previous speculation that CO_2 emission arose continuously (deriv > 0 in log scale), while its rate of change decreases substantially. Similarly, we narrow down our focus to the economic downturn caused by COVID-19, in which sample derivatives are positive as well, but still difficult to verify the change of CO_2 growth rate given the insufficient data within such a short window. However, it is likely to see the shallow trend in the near future, based on the model prediction.

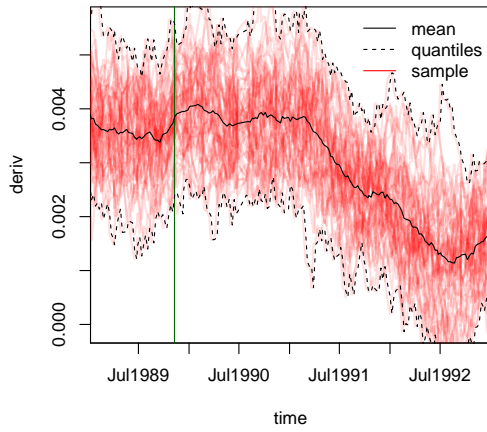Table 1: Random effects posterior means and 95% credible intervals in terms of relative rates

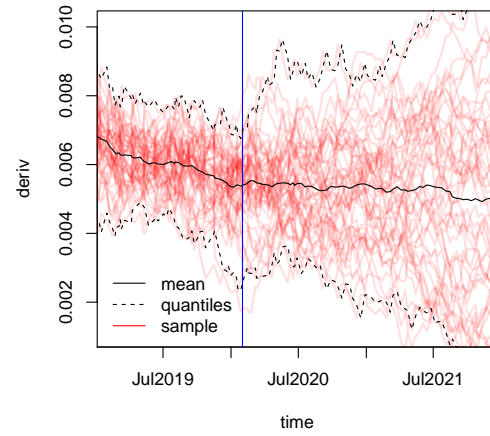|	0.5quant	0.025quant	0.975quant
sd for gamma	1.000	1.000	1.000
sd for timeInla	1.002	1.002	1.002

(a) Fitted values
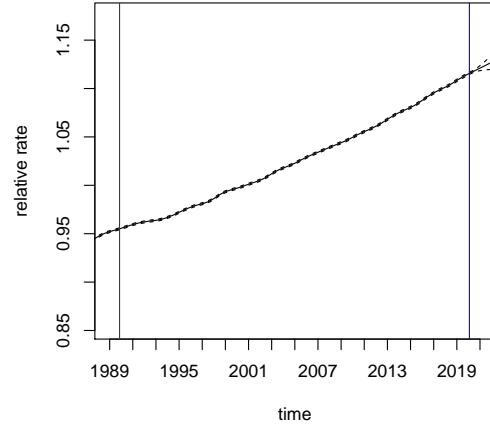
(b) Trend of random effect

(c) Sample derivatives during 1989-1992

(d) Sample derivatives during 2019-2021

Figure 1: CO2 model results

Conclusion

In summary, we investigated the impact of two events of interest on global CO_2 concentrations. Although the sample derivatives of our model show positive sign that CO_2 increased after both events, they are all decreasing and that the corresponding trends become much shallower, especially the industrialization change alongside the Fall of the Berlin Wall. Thus, we conclude that the Fall of the Berlin Wall slowed down CO_2 emission, yet the impact of COVID-19 is currently insignificant.

Appendix

```r
## Question 1 ##
rm(list=ls())

cUrl=paste0("http://scrippsco2.ucsd.edu/assets/data/atmospheric/",
            "stations/flask_co2/daily/daily_flask_co2_mlo.csv")
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$quality==0, ]
co2s$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(prec=list(prior='pc.prec',param=c(0.1,0.5)))),
            control.inla=list(strategy='gaussian'),
            control.predictor=list(compute=TRUE,link=1),
            control.compute=list(config=TRUE),
            verbose=FALSE)

sampleList=INLA::inla.posterior.sample(50, co2res,selection=list(timeInla=0))
sampleMean=do.call(cbind, Biobase::subListExtract(sampleList,"latent"))
sampleDeriv=apply(sampleMean,2, diff)/diff(co2res$summary.random$timeInla$ID)

qCols=c('0.5quant','0.025quant','0.975quant')
table <- round(exp(Pmisc::priorPost(co2res)$summary[,qCols]), 3)
kable(table,
      booktabs = TRUE,
      caption = "Random effects posterior means and 95%
      credible intervals in terms of relative rates")

# F1
matplot(co2ext$day, (co2res$summary.fitted.values[,qCols]),
        type ="l",col ="black",lty =c(1,2,2),log ="y",xlab ="time",ylab="ppm",
        xlim=c(as.Date("1989-01-01"), as.Date("2021-01-01")), xaxt='n',)
```

```r
forX=as.Date(c("1989-01-01","2021-01-01"))
forX=seq(forX[1], forX[2],by ="24 months")
axis(1,as.numeric(forX),format(forX,"%Y"))
abline(v = as.Date("1989-11-09"), col = "dark green")
abline(v = as.Date("2020-02-01"), col = "blue")


# F2
sampleDeriv=apply(sampleMean,2, diff)/diff(co2res$summary.random$timeInla$ID)

Stime=timeOrigin+round(365.25*co2res$summary.random$timeInla$ID)
cset1 = GET::create_curve_set(list(r = as.numeric(Stime[-1]),obs = sampleDeriv))
myEnv1 = GET::central_region(cset1, coverage = 0.95)

matplot(Stime, exp(co2res$summary.random$timeInla[, qCols]),
        type ="l",col ="black",lty =c(1,2,2),xlab ="time",ylab ="relative rate",
        xlim=c(as.Date("1989-01-01"), as.Date("2021-01-01")), xaxt='n')
forX=as.Date(c("1989-01-01","2021-01-01"))
forX=seq(forX[1], forX[2],by ="24 months")
axis(1,as.numeric(forX),format(forX,"%Y"))
abline(v = as.Date("1989-11-09"), col = "dark green")
abline(v = as.Date("2020-02-01"), col = "blue")


# F3 and 4
forX=as.Date(c("1989-01-01","1993-01-01"))
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 ="#FF000020",
        xlab ="time",ylab ="deriv",
        xaxt ="n",ylim =quantile(sampleDeriv[toPlot,],c(0.01,0.995)))
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("1989-11-09"), col = "dark green")
matlines(Stime[-1], as.data.frame(myEnv1)[,c("lo","hi","central")],
        lty = c(2,2,1),col = "black")
legend("topright",bty ="n",lty =c(1,2,1), col =c("black","black","red"),
        legend =c("mean","quantiles","sample"))


forX=as.Date(c("2019-01-01","2022-01-01"))
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 ="#FF000020",
        xlab ="time",ylab ="deriv",
        xaxt ="n",ylim =quantile(sampleDeriv[toPlot,],c(0.01,0.995)))
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("2020-02-01"), col = "blue")
matlines(Stime[-1], as.data.frame(myEnv1)[,c("lo","hi","central")],
        lty = c(2,2,1),col = "black")
legend("bottomleft",bty ="n",lty =c(1,2,1), col =c("black","black","red"),
        legend =c("mean","quantiles","sample"))
```

COVID-19 Mortality Analysis

Introduction

The COVID-19 epidemic does not only impose huge economic loss globally, but it also cause a large number of deaths. This report studies the weekly mortality data in Quebec, with respect to the COVID-19 influence. In particular, we first evaluate the exposure trend and excess deaths, considering different age groups of patients. Using Bayesian inference and semi-parametric time trend analysis, we examine the research hypothesis that the deaths of the elderly were comparably higher than previous in the first wave (March to May), but in the corresponding period, young people showed indifferent results with the historical benchmarks. The situation was also hypothesized to overturn between the two age groups in the second wave (September to recent weeks).

Method

We assume that the response *Death* follows the Poisson distribution, since death counts are positive discrete. A generalized additive model (GAM) with a log link is proposed below to interpret the data, as follows:

$$Y_i \sim \text{Poisson}\left(O_i \lambda_i\right)$$
$$\log\left(\lambda_i\right) = X_i \beta + U\left(t_i\right) + V_i$$
$$V_i \sim \text{i.i.d Normal}(0, \sigma_V^2)$$
$$[U_1 \dots U_T]^T \sim \text{RW2}\left(0, \sigma_U^2\right)$$
$$X_{i0} = 1$$
$$X_{i1} = \cos\left(2\pi t_i / 365.25\right)$$
$$X_{i2} = \sin\left(2\pi t_i / 365.25\right)$$
$$X_{i3} = \cos\left(2\pi t_i / 182.625\right)$$
$$X_{i4} = \sin\left(2\pi t_i / 182.625\right)$$

where we use frequencies of 6 and 12 months in modelling. X_{i1}, X_{i2}, X_{i3} and X_{i4} are four sinusoidal basis functions to take account for a wide range of seasonal effects on deaths, and their coefficient estimates are expressed as β.

In addition, $U(t)$ is a second-order random walk with parameter σ_U^2, while V_i is identically independent distributed random variables that follow i.i.d Normal$(0, \sigma_V^2)$. Using Bayesian inference, we can choose the hyperparameters σ_u, and σ_v alongside some prior knowledge in a sensible approach. It is common that the random slope, which is the log rate of change in the number of deaths, could vary in a range of 0.01 from time to time. We also consider a 20% difference of death counts varied at each measurement. Therefore, we set all these hyperparameters using penalized complexity priors for precision in R-INLA, as follow:

$$P(\sigma_V^2 > \log(1.2)) = 0.5 \Longrightarrow \sigma_V^2 \sim \exp\left(\log(0.8)\right)$$
$$P(\sigma_U^2 > 0.01) = 0.5 \Longrightarrow \sigma_U^2 \sim \exp\left(100\log(2)\right)$$

With such time series data of mortality, the use of first harmonics is beneficial to handle the periodicity. Thus, the proposed model does not only include seasonal effects, but it also involves natural variation thorough its underlying random effect and random slope terms.

Results

We here fit two separate models with the data of different age groups: under 50's and over 70's. To better observe death counts at different time points, black and green vertical lines mark the start of March and September, respectively. In this section, we help to answer how different populations were affected by this global epidemic.

1. Under 50's Deaths were consistent with previous years in March, but increased in recent weeks?
Table 2 delivers a brief model summary in terms of fixed and random effects in a natural scale. Particularly, the intercept of 62.085 indicates the average death counts of under 50's. The relative rate is expected to be 1.044 among data, and the growth rate is typically 1.013 different between data points. To interpret, it is not surprising

to see if the death counts go up by 1.3% from week to week, and there could be around 4.4% fluctuation by natural randomness even in the same week

We present a few figures to better assess the research hypotheses. Figure 2(a) illustrates the proposed linear model predicting the mortality data under 50's where we cannot find significant discrepancy when comparing the two waves to previous years. In fact, the model cannot explain the group of under 50's very well, since there are always a few outliers from time to time. Both time trends seem to have high variation around the two lines, as highlighted by the Figure 2(b). The general trend is negative and shallow, indicating the decrease in death number of under 50's over the past few years. This might be intuitively explained by better health care services nowadays.

For more rigorous evaluation, we compare the true data with the posterior quantiles in Figure 2(c); true data are well captured by the 95% coverage. In addition, the pattern of excess death under 50's resembles some white noise in Figure 2(d), so we think their death counts are normal. The total numbers of excess deaths under 50's are quite closed in the two periods.

Table 2: Effects posterior means and 95% credible intervals for under 50's death

|	0.5quant	0.025quant	0.975quant
(Intercept)	62.085	61.335	62.838
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: Cumulative excess death for under 50's from 2020/03/01 to 2020/06/01
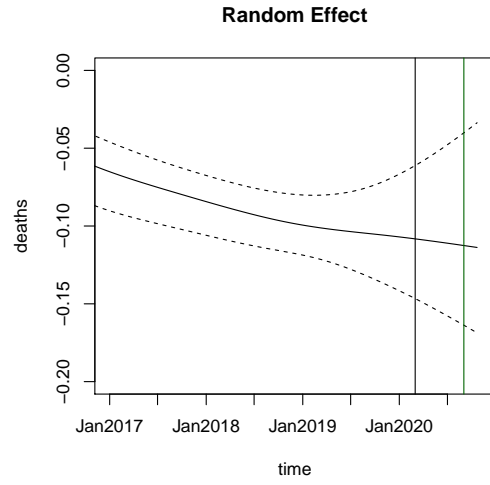
0%	25%	50%	75%	100%
37	75	100	114	174

Table 4: Cumulative excess death for under 50's from 2020/09/01 to 2020/11/01

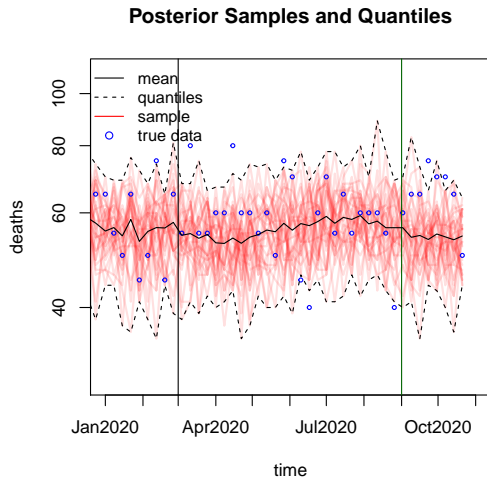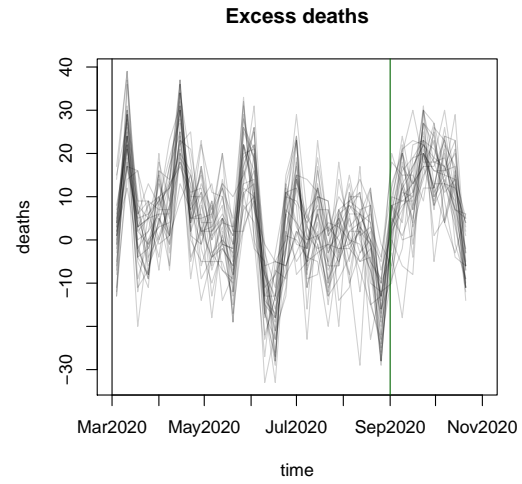| 0%	25%	50%	75%	100%
36	63	96	102	124
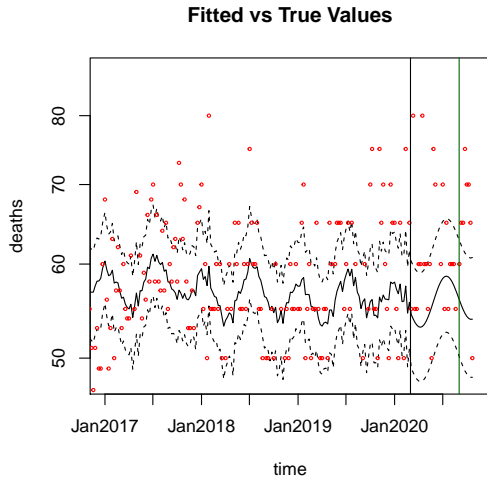
(a) Fitted values

(b) Trend of random effect

(c) Posterior samples and quantiles

(d) Excess death

Figure 2: Model results for under 50's

2. Deaths amongst the elderly were well above the historical averages in spring, but indifferent in recent weeks?

Table 5 shares a brief model summary in terms of fixed and random effects in a natural scale. Particularly, the intercept of 884.655 indicates the average death counts of the elderly. The relative rate is expected to be 1.041 among data, the growth rate is typically 1.174 different from each data point. That being said, we may see 17.4% difference in elderly death counts from week to week, which is much larger than young people. Within 4.1% normal fluctuation is reliable at the same data point when predicting.

Several figures are presented below to better investigate the COVID impact on the population of the elderly. Figure 3(a) illustrates the corresponding linear fit predicting the mortality data over 70's, in which some true data seem to scatter above the historical average from March to May, whereas the model looks capable of capturing the data from September to now. In the second wave, it is worth-noting that most of the true data are above the median.

According to Figure 3(b), a downward trend with higher variation appears similarly with the under 50's. Although Figure 3(c) demonstrates the broad coverage of our model against the true death counts, we still inspect many excess deaths from the two waves in Figure 3(d). In May, the true deaths are much more than our model prediction. Subsequently, the trend of excess death for over 70's goes down, but arises again in more recent weeks. As of September, the number of over 70's deaths has reached 2773 until recently. Nevertheless, the weekly number of excess deaths is merely above zero level a little. Fewer than 200 deaths are found, but Quebec is a big province, and the elderly themselves are more vulnerable. Hence, we feel less confident to say the recent number of deaths goes up significantly.

Table 5: Effects posterior means and 95% credible intervals for over 70's death

	0.5quant	0.025quant	0.975quant
(Intercept)	884.655	877.944	891.148
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 6: Cumulative excess death for over 70's from 2020/03/01 to 2020/06/01
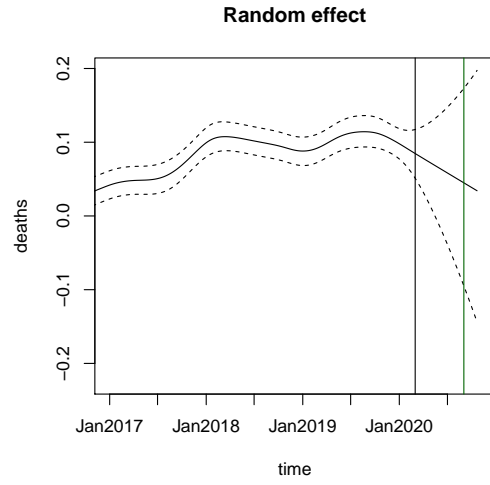
0%	25%	50%	75%	100%
3471	4404	4566	4847	5141

Table 7: Cumulative excess death for over 70's from 2020/09/01 to 2020/11/01

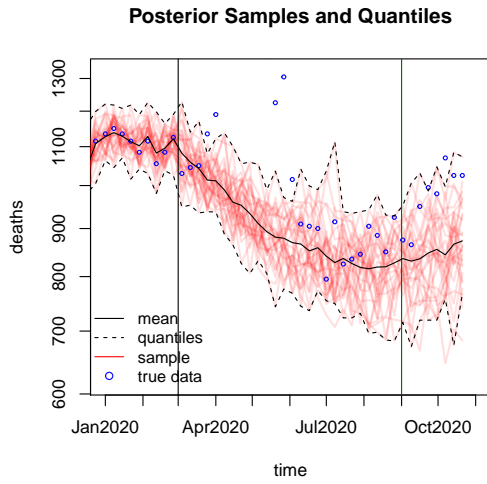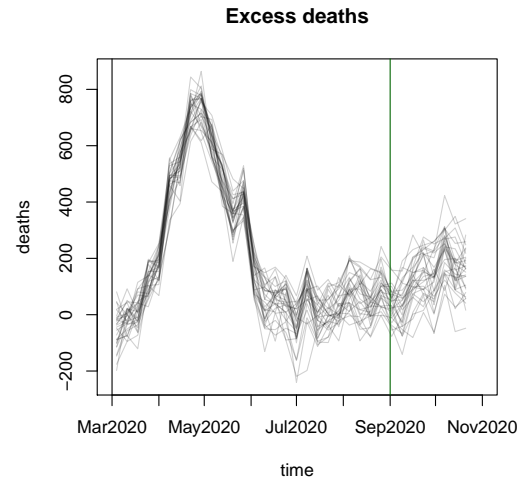| 0%	25%	50%	75%	100%
-113	658	1036	1359	2121
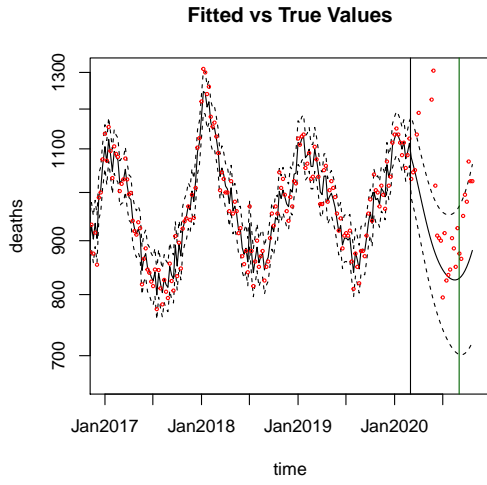
(a) Fitted values

(b) Trend of random effect

(c) Posterior samples and quantiles

(d) Excess death

Figure 3: Model results for over 70's

Conclusion

In general, this report contributes to better understanding of the effect of COVID-19 on different age groups. Using Bayesian inference on top of a semi-parametric model, we first find the death counts of under 50's are relatively stable during the two waves; their mortality data has been in line over the course. Second, death numbers are unfriendly to the elderly. Deaths of over 70's in Quebec exceed the historical average tremendously in the spring. However, their excess deaths in the second wave are captured by the 95% credible intervals of our model, so we think deaths are not significantly more than would be expected pre-covid.

Appendix

```r
## Question 2 ##
rm(list=ls())
x <- read.csv('death.csv')
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, ]

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),]

xForInla$timeNumeric = as.numeric(xForInla$time)
xForInla$timeForInla = (xForInla$timeNumeric -
                          as.numeric(as.Date("2015/1/1")))/365.25
xForInla$timeIid = xForInla$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)

# Under 50's
xForInlaTotal=xForInla[xForInla$age=='0-49 years old', ]

res=inla(dead~sin12+sin6+cos12+cos6+
           f(timeIid,prior='pc.prec',param=c(log(1.2),0.5))+
           f(timeForInla,model ='rw2',prior='pc.prec',param=c(0.01,0.5)),
         data=xForInlaTotal,
         control.inla = list(fast=FALSE, strategy='laplace'),
         control.predictor =list(compute=TRUE,link=1),
         control.compute =list(config=TRUE),
         family='poisson')

qCols=paste0(c(0.5,0.025,0.975),"quant")
table <- rbind(res$summary.fixed[, qCols], Pmisc::priorPostSd(res)$summary[,qCols])
kable(round(exp(table),3),
      booktabs = TRUE,
      caption = "Effects posterior means and 95%
      credible intervals for under 50's death")

sampleList=INLA::inla.posterior.sample(30, res,selection =list(Predictor =0))
sampleIntensity=exp(do.call(cbind, Biobase::subListExtract(sampleList,"latent")))
sampleDeaths=matrix(rpois(length(sampleIntensity),sampleIntensity),
                    nrow(sampleIntensity),ncol(sampleIntensity))
```

```r
cset1 = GET::create_curve_set(list(r = as.numeric(xForInlaTotal$time),obs =sampleDeaths))
myEnv1 = GET::central_region(cset1, coverage = 0.95)


# F1 Fitted values and real data
matplot(xForInlaTotal$time, res$summary.fitted.values[,qCols],
        type ="l",lty =c(1,2,2),col ="black",log ="y",
        xlim=c(as.Date("2017-01-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Fitted vs True Values")
points(x[x$age=="0-49 years old",c("time","dead")],cex =0.4,col ="red",)
forX=as.Date(c("2017-01-01","2020-11-01"))
forX=seq(forX[1], forX[2], by ="6 month")
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")


# F2 Random effects
matplot(xForInlaTotal$time,
        res$summary.random$timeForInla[,c("0.5quant","0.975quant","0.025quant")],
        type ="l",lty =c(1,2,2),col ="black",ylim = c(-0.2, 0),
        xlim=c(as.Date("2017-01-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Random Effect")
forX=as.Date(c("2017-01-01","2020-11-01"))
forX=seq(forX[1], forX[2],by ="6 months")
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")


# F3 Posterior data and real data
matplot(xForInlaTotal$time, sampleDeaths,col ="#FF000120",lwd =2,lty =1,type ="l",log ="y",
        xlim=c(as.Date("2020-01-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Posterior Samples and Quantiles")

points(x[x$age=="0-49 years old",c("time","dead")],cex =0.5,col ="blue",)
forX=as.Date(c("2020-01-01","2020-11-01"))
forX=seq(forX[1], forX[2],by ="1 months")
axis(1,as.numeric(forX),format(forX,"%b%Y"))

abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")

matlines(xForInlaTotal$time,
         as.data.frame(myEnv1)[,c("lo","hi","central")],
         lty = c(2,2,1),
         col = "black")
legend("topleft",bty ="n",
       lty =c(1,2,1,NA),
       pch = c(NA, NA, NA, 1),
       col =c("black","black","red","blue"),
       legend =c("mean","quantiles","sample", "true data"),
       cex = 0.9)
```

```r
# F4 Excess death
xPostCovidTotal=xPostCovid[xPostCovid$age=="0-49 years old",]
xPostCovidForecast=sampleDeaths[match(xPostCovidTotal$time,xForInlaTotal$time), ]
excessDeaths=xPostCovidTotal$dead-xPostCovidForecast

matplot(na.omit(xPostCovidTotal$time), excessDeaths,
        type = "l",lty = 1, col = "#00000030",
        xlim=c(as.Date("2020-03-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Excess deaths")
forX=as.Date(c("2020-03-01","2020-11-01"))
forX=seq(forX[1], forX[2],by ="1 months")
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")


excessDeathsSub=excessDeaths[xPostCovidTotal$time>as.Date("2020/03/01")
                             &xPostCovidTotal$time<as.Date("2020/06/01"), ]
excessDeathsInPeriod=apply(excessDeathsSub,2, sum)
kable(rbind(round(quantile(excessDeathsInPeriod))),
      booktabs = TRUE,
      caption = "Cumulative excess death for under 50's from 2020/03/01 to 2020/06/01")

excessDeathsSub=excessDeaths[xPostCovidTotal$time>as.Date("2020/09/01")
                             &xPostCovidTotal$time<as.Date("2020/11/01"), ]
excessDeathsInPeriod=apply(excessDeathsSub,2, sum)
kable(rbind(round(quantile(excessDeathsInPeriod))),
      booktabs = TRUE,
      caption = "Cumulative excess death for under 50's from 2020/09/01 to 2020/11/01")

#Over 70's
xForInlaTotal=xForInla[xForInla$age=='70 years old and over', ]

res=inla(dead~sin12+sin6+cos12+cos6+
             f(timeIid,prior='pc.prec',param=c(log(1.2),0.5))+
             f(timeForInla,model ='rw2',prior='pc.prec',param=c(0.01,0.5)),
         data=xForInlaTotal,
         control.inla = list(fast=FALSE, strategy='laplace'),
         control.predictor =list(compute=TRUE,link=1),
         control.compute =list(config=TRUE),
         family='poisson')

qCols=paste0(c(0.5,0.025,0.975),"quant")
table <- rbind(res$summary.fixed[, qCols], Pmisc::priorPostSd(res)$summary[,qCols])
kable(round(exp(table),3),
      booktabs = TRUE,
      caption = "Effects posterior means and 95%
      credible intervals for over 70's death")

sampleList=INLA::inla.posterior.sample(30, res,selection =list(Predictor =0))
sampleIntensity=exp(do.call(cbind, Biobase::subListExtract(sampleList,"latent")))
sampleDeaths=matrix(rpois(length(sampleIntensity),sampleIntensity),
                    nrow(sampleIntensity),ncol(sampleIntensity))


cset1 = GET::create_curve_set(list(r = as.numeric(xForInlaTotal$time),obs =sampleDeaths))
myEnv1 = GET::central_region(cset1, coverage = 0.95)
```

```r
# F1 Fitted values and real data
matplot(xForInlaTotal$time, res$summary.fitted.values[,qCols],
        type ="l",lty =c(1,2,2),col ="black",log ="y",
        xlim=c(as.Date("2017-01-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Fitted vs True Values")
points(x[x$age=="70 years old and over",c("time","dead")],cex =0.4,col ="red",)
forX=as.Date(c("2017-01-01","2020-11-01"))
forX=seq(forX[1], forX[2], by ="6 month")
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")


# F2 Random effects
matplot(xForInlaTotal$time,
        res$summary.random$timeForInla[,c("0.5quant","0.975quant","0.025quant")],
        type ="l",lty =c(1,2,2),col ="black",
        xlim=c(as.Date("2017-01-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Random effect")
forX=as.Date(c("2017-01-01","2020-11-01"))
forX=seq(forX[1], forX[2],by ="6 months")
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")


# F3 Posterior data and real data
matplot(xForInlaTotal$time, sampleDeaths,col ="#FF000120",lwd =2,lty =1,type ="l",log ="y",
        xlim=c(as.Date("2020-01-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Posterior Samples and Quantiles")

points(x[x$age=="70 years old and over",c("time","dead")],cex =0.5,col ="blue",)
forX=as.Date(c("2020-01-01","2020-11-01"))
forX=seq(forX[1], forX[2],by ="1 months")
axis(1,as.numeric(forX),format(forX,"%b%Y"))

abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")

matlines(xForInlaTotal$time,
         as.data.frame(myEnv1)[,c("lo","hi","central")],
         lty = c(2,2,1),
         col = "black")
legend("bottomleft",bty ="n",
       lty =c(1,2,1,NA),
       pch = c(NA, NA, NA, 1),
       col =c("black","black","red","blue"),
       legend =c("mean","quantiles","sample", "true data"),
       cex = 0.9)


# F4 Excess death
xPostCovidTotal=xPostCovid[xPostCovid$age=="70 years old and over",]
xPostCovidForecast=sampleDeaths[match(xPostCovidTotal$time,xForInlaTotal$time), ]
excessDeaths=xPostCovidTotal$dead-xPostCovidForecast
```

```r
matplot(na.omit(xPostCovidTotal$time), excessDeaths,
        type = "l",lty = 1, col = "#00000030",
        xlim=c(as.Date("2020-03-01"), as.Date("2020-11-01")), xaxt='n',
        xlab = "time", ylab="deaths",
        main="Excess deaths")
forX=as.Date(c("2020-03-01","2020-11-01"))
forX=seq(forX[1], forX[2],by ="1 months")
axis(1,as.numeric(forX),format(forX,"%b%Y"))
abline(v = as.Date("2020-03-01"), col = "black")
abline(v = as.Date("2020-09-01"), col = "dark green")
```

```r
excessDeathsSub=excessDeaths[xPostCovidTotal$time>as.Date("2020/03/01")
                            &xPostCovidTotal$time<as.Date("2020/06/01"), ]
excessDeathsInPeriod=apply(excessDeathsSub,2, sum)
kable(rbind(round(quantile(excessDeathsInPeriod))),
      booktabs = TRUE,
      caption = "Cumulative excess death for over 70's from 2020/03/01 to 2020/06/01")

excessDeathsSub=excessDeaths[xPostCovidTotal$time>as.Date("2020/09/01")
                            &xPostCovidTotal$time<as.Date("2020/11/01"), ]
excessDeathsInPeriod=apply(excessDeathsSub,2, sum)
kable(rbind(round(quantile(excessDeathsInPeriod))),
      booktabs = TRUE,
      caption = "Cumulative excess death for over 70's from 2020/09/01 to 2020/11/01")
```