Aalto University
School of Science

Project

CS-E5710: Bayesian Data Analysis

Anonymous1 XXXXXX
Anonymous2 XXXXXX
Anonymous3 XXXXXX

Returned: December 17, 2021

Contents

1 Introduction

Life expectancy at birth is a statistical measure of average number of years a newborn is expected to live based on age-specific mortality rates in a given period. It is still worth noting that the actual age-specific death rate of any specific birth cohort cannot be known in advance. Life expectancy assesses population health and can be used to predict how a population ages. Life expectancy in the world has been growing quite steadily for at least since the 1950s. Being able to predict how a population ages and grows is crucial to be able to prepare for changes that for example ageing population causes in the society.

This presentation tries to answer to the question of how life expectancy in the world develops until 2050. It is well known fact that life expectancy has been going up in all over the world due to advancements in medical science and improvements in quality of life. Even though life expectancy has increased steadily the maximum lifespan of a human is still from 1997 when Jeanne Calment died at the age of 122. According to Guinness World Records she still holds this record and currently oldest person living is 118 years old. This would suggest that there is indeed a maximum lifespan and the increase in life expectancy should end or at least slow down at some point in the future. On the other hand according to United Nations the amount of centenarians is exponentially growing which could suggest that it would be reasonable to expect life expectancy to keep increasing at some rate until 2050[2].

The main idea for the modeling is to predict the life expectancy in Europe until 2050 using three models. The models used were linear model, cubic model and exponential model. Life expectancy has increased in most areas almost linearly as seen in figure 1. However, as life expectancy is mostly quite linear at this point in time linear model makes sense in predicting life expectancy. In Europe the life expectancy curve has periods of faster growth and slower growth but overall the trend is quite linear. In certain countries there has been different kind of crises such as widely spread AIDS in sub-Saharan Africa that caused life expectancy to drop significantly at certain times. As these kinds of drops are nearly impossible to predict it was decided not to predict these areas. The exponential model is a model that converges towards a certain value which is in line with the expectation of maximum human life span. However, as life expectancy in Europe had periods of faster growth and more stagnant periods of life expectancy cubic model was chosen as it could be able to model these time periods if they are somewhat cyclical.

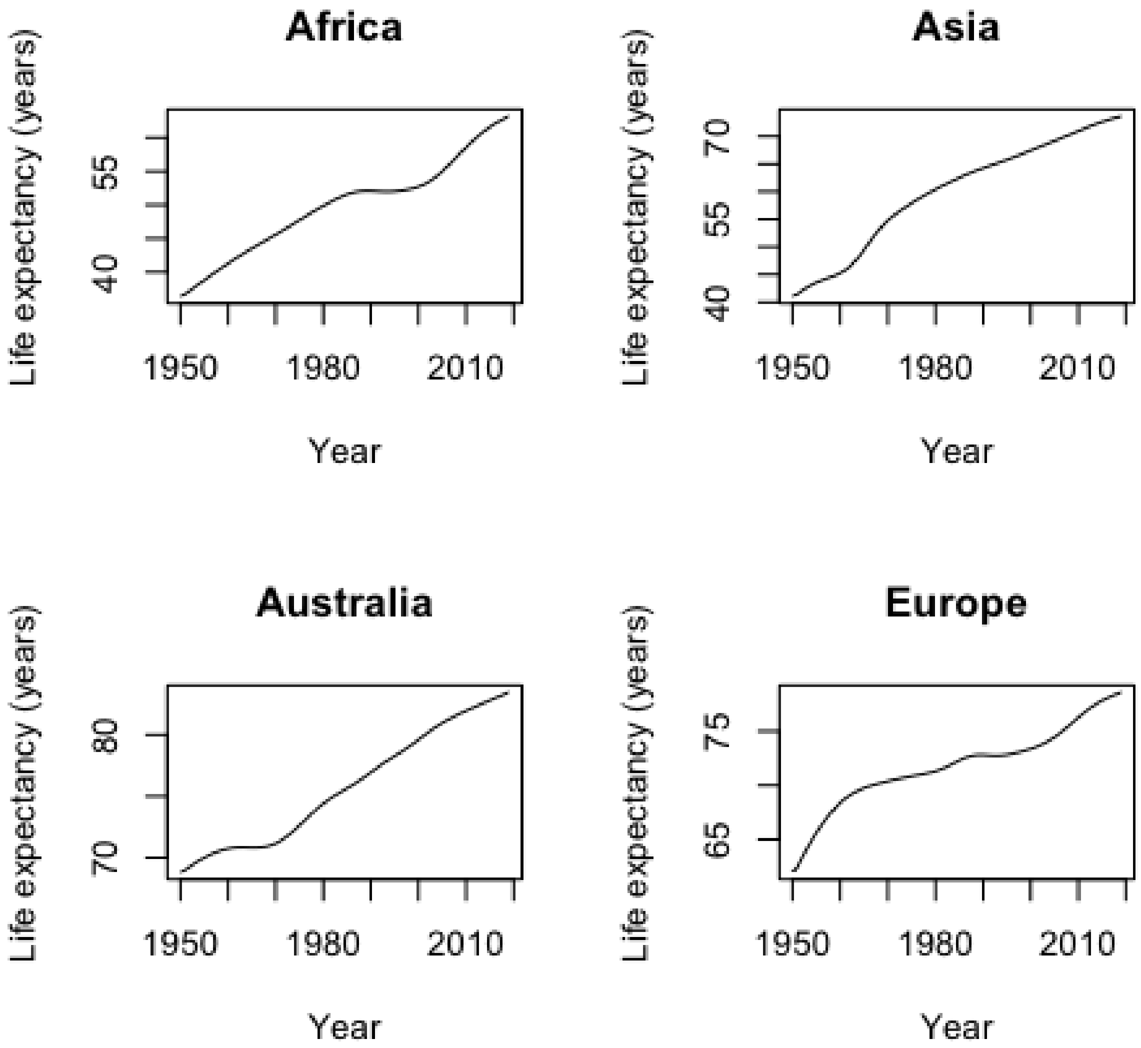

Figure 1: Life expectancy in Africa, Asia, Australia and Europe from 1950 to 2020.

The report structure is the following. In part 2 the data and analysis of the problem is briefly explained, in part 3 the models and priors are described, in part 4 the Stan model options and results are shown, part 5 is discussion and conclusions and finally part 6 is self-reflection. The code used is shown in the appendix.

2 Data and analysis of the problem

The data used in this analysis is life expectancy at birth data for all countries, continents and the world obtained from [3]. The goal of this report is to answer to how life expectancy in all parts of the world will change until 2050. This particular data hasn't been used in an online case study that we could find. However, The United Nations makes World Population Prospects every two years which includes probabilistic projections until 2100 for life expectancy in all nations, continents and world[4, 5]. Even though The United Nations uses different data, the idea of the analysis is similar. They made the projections using Bayesian data analysis and more specifically hierarchical model. The main differences between these projections are the timeframe (from present time to 2100 vs 2050), models and the dataset used. Still the goal of these analyses is very similar. ´

3 Models

All models have Normal distributed residuals, i.e. errors between model means and observations. This can be represented by the following model

$$y_j(x) \sim \text{Normal}(\mu_j(x), \sigma), \tag{1}$$

where $y_j(x)$ is the observation and $\mu_j(x)$ is the model mean of the j:th group on year x. Parameter $\sigma > 0$ is the standard error of the residuals that will be fitted in the model. This parameter was given the following weakly informative common prior in all models

$$\sigma \sim \text{Inv-}\chi^2(0.01) = \text{Scaled-Inv-}\chi^2(0.01, \frac{1}{0.01}), \tag{2}$$

which describes a distribution with a standard deviation of $\sqrt{100} = 10$ [1]. This seems reasonable choice as the observed data is in the same order of magnitude (roughly 40-80).

For all models the x-data was shifted to 1-70 instead of the years 1950-2020. This was done to help choosing suitable priors and also to improve numerical stability for the cubic model. In the following subsections only the dependence between model means $\mu_j(x)$ and predictors x is discussed as the part presented in this section is same for all models.

3.1 Linear model

The linear model without priors is described by

$$\begin{aligned} &\mu_j(x) = a_j + b_j x \\ &a \sim \text{Normal}(\mu_a, \sigma_a) \\ &b \sim \text{Normal}(\mu_b, \sigma_b), \end{aligned} \tag{3}$$

where μ_b, μ_a, σ_b are σ_a are free hyperparameters. The model was fitted with the following prior distributions for these parameters

$$
\begin{aligned}
\mu_a &\sim \text{Normal}(50, 20) \\
\mu_b &\sim \text{Normal}(0, 10) \\
\sigma_a &\sim \text{Inv-}\chi^2(0.01) \\
\sigma_b &\sim \text{Inv-}\chi^2(0.01).
\end{aligned} \tag{4}
$$

The intercept mean parameter μ_a was given an (weakly) informative prior as we know that in year 1950 its value must be somewhere in the range of realistic life-expectancies 30-70. For the slope mean parameter μ_b we can certainly expect that life expectancy is not increasing or decreasing over 10 years in just one year in the long run. The standard error priors σ_a and σ_b were given a similar inverse-chi-squared prior as for the regression standard error giving the hyperparameters an expected prior deviation of 10.

3.2 Cubic model

The cubic model without priors is described by

$$
\begin{aligned}
&\mu_j(x) = a_j x^3 + b_j x^2 + c_j x + d_j \\
&a \sim \text{Normal}(\mu_a, \sigma_a) \\
&b \sim \text{Normal}(\mu_b, \sigma_b) \\
&c \sim \text{Normal}(\mu_c, \sigma_c) \\
&d \sim \text{Normal}(\mu_d, \sigma_d),
\end{aligned} \tag{5}
$$

The priors used for this model were

$$
\begin{aligned}
\mu_a &\sim \text{Normal}(0, 0.1) \\
\mu_b &\sim \text{Normal}(0, 1) \\
\mu_c &\sim \text{Normal}(0, 10) \\
\mu_d &\sim \text{Normal}(0, 100) \\
\sigma_a &\sim \text{Inv-}\chi^2(10) \\
\sigma_b &\sim \text{Inv-}\chi^2(1) \\
\sigma_c &\sim \text{Inv-}\chi^2(0.1) \\
\sigma_d &\sim \text{Inv-}\chi^2(0.01)
\end{aligned} \tag{6}
$$

The logic behind these values is that the cubic term for example becomes very large towards the end of the time series $x^3 = 70^3 = 343000$ so we can definitely expect the prefactor for this term to be much smaller than for the the other terms with lower powers. The same idea is applied to both the Normal priors for the means and the inverse-chi-squared priors for the standard deviations.

3.3 Exponential model

The exponential model without priors is described by

$$
\begin{aligned}
\mu_j(x) &= a_j - b_j \lambda_j^x \\
a &\sim \text{Normal}(\mu_a, \sigma_a) \\
b &\sim \text{Normal}(\mu_b, \sigma_b) \\
\lambda &\sim \text{Beta}(\alpha, \beta),
\end{aligned}
\tag{7}
$$

Beta distribution was chosen for λ as we want to restrict it to have values between 0 and 1. Values larger than 1 would result in exponential growth, which we assume to be impossible in the long run for life expectancy and values below 0 would result in negative and/or oscillating values which is simply not possible. In practise the parameters β and α are also positive as we are expecting a curve approaching the positive limit α from below, not above.

The priors used were

$$
\begin{aligned}
\mu_a &\sim \text{Normal}(0, 100) \\
\mu_b &\sim \text{Normal}(0, 100) \\
\sigma_a &\sim \text{Inv-}\chi^2(0.01) \\
\sigma_b &\sim \text{Inv-}\chi^2(0.01) \\
\alpha &\sim \text{Exp}(0.1) \\
\beta &\sim \text{Exp}(0.1).
\end{aligned}
\tag{8}
$$

The mean parameters were given quite weakly informative Normal priors around 0. The deviations were given the same priors as for the linear case and beta distribution α and β parameters were given exponential priors with a mean 10. The exponential priors imply that α and β draws have the highest density being close to 0, meaning that the Beta distribution is in most cases quite wide only going down close to 0 and 1.

4 Results

4.1 Running the stan model

All stan models were run with following R commands

```
model_i <- rstan::stan_model(file = "stanfiles/model_code_i.stan")

fit_i <- rstan::sampling(model_i, data = data_continents,
control = list(max_treedepth = 15)),
```

where i $\in$ {linear, cubic, exponential} corresponds to the current model and "data$_c$ontinents". The only non-default option used for all of the models was increasing the simulation treedepth from 10 to 15. This was done due to a portion of the iterations exceeding the default treedepth with all models. For the cubic model this didn't fully fix the problem, but we decided not to increase the treedepth anymore as the model already took very long time to run.

4.2 Predictions

The predicted means and 90% predictive intervals until the year 2050 are shown in figures: Fig. 2 for linear, Fig. 3 for cubic and Fig. 4 for exponential model. As we can see from these figures, linear and exponential models predict that life expectancy in Europe will keep rising until 2050 with high probability while the cubic model predicts that life expectancy will start to decrease rapidly around 2020. however, the cubic model prediction has very large uncertainty after 2020 and 5% of the draws indicate a still increasing life expectancy in year 2050. Only predictions for Europe are visualised in these figures, even though in total, the data from 9 continents were used to fit the hierarchical models.

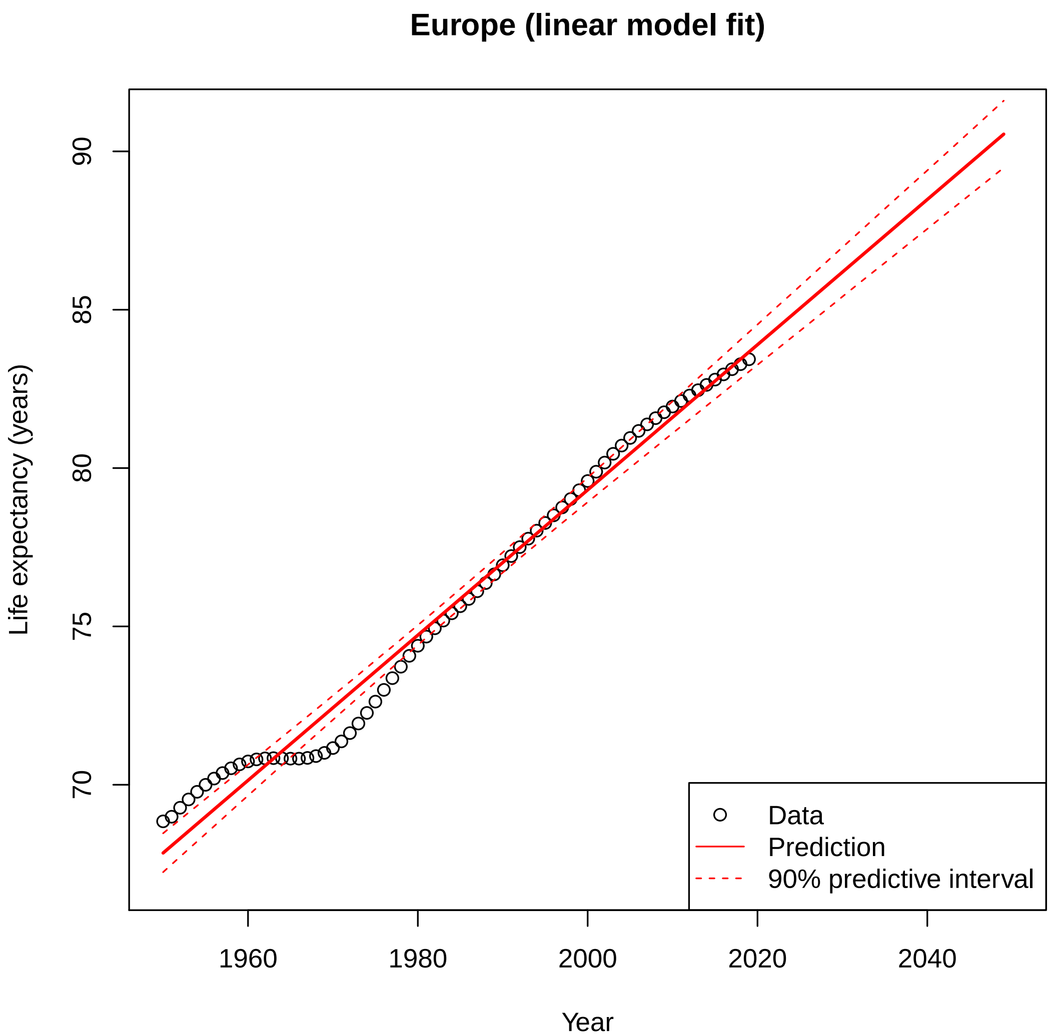

Figure 2: Data, predicted model means, and 90% predictive intervals of life expectancy in Europe between years 1950-2050 using the linear model (Eq. 3).

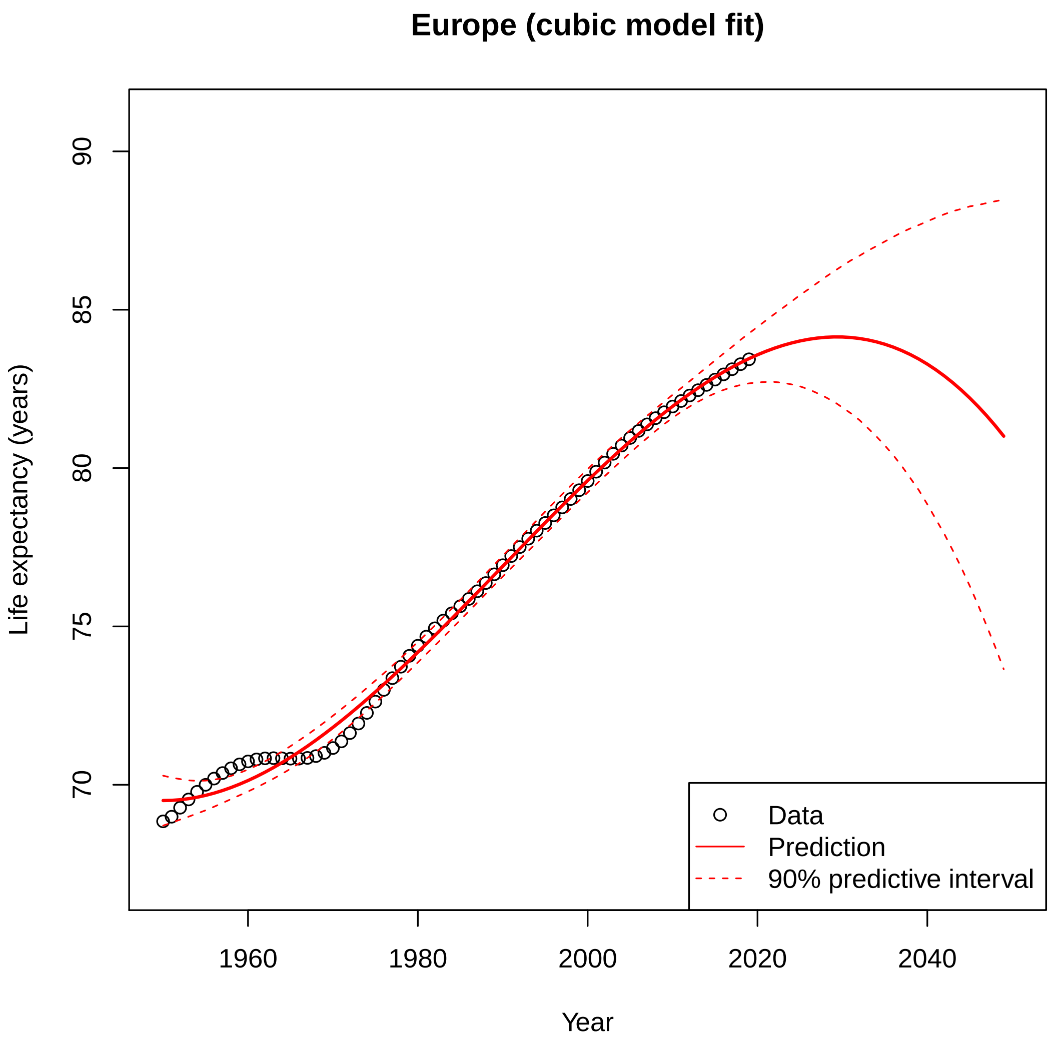

Figure 3: Data, predicted model means, and 90% predictive intervals of life expectancy in Europe between years 1950-2050 using the cubic model (Eq. 5).

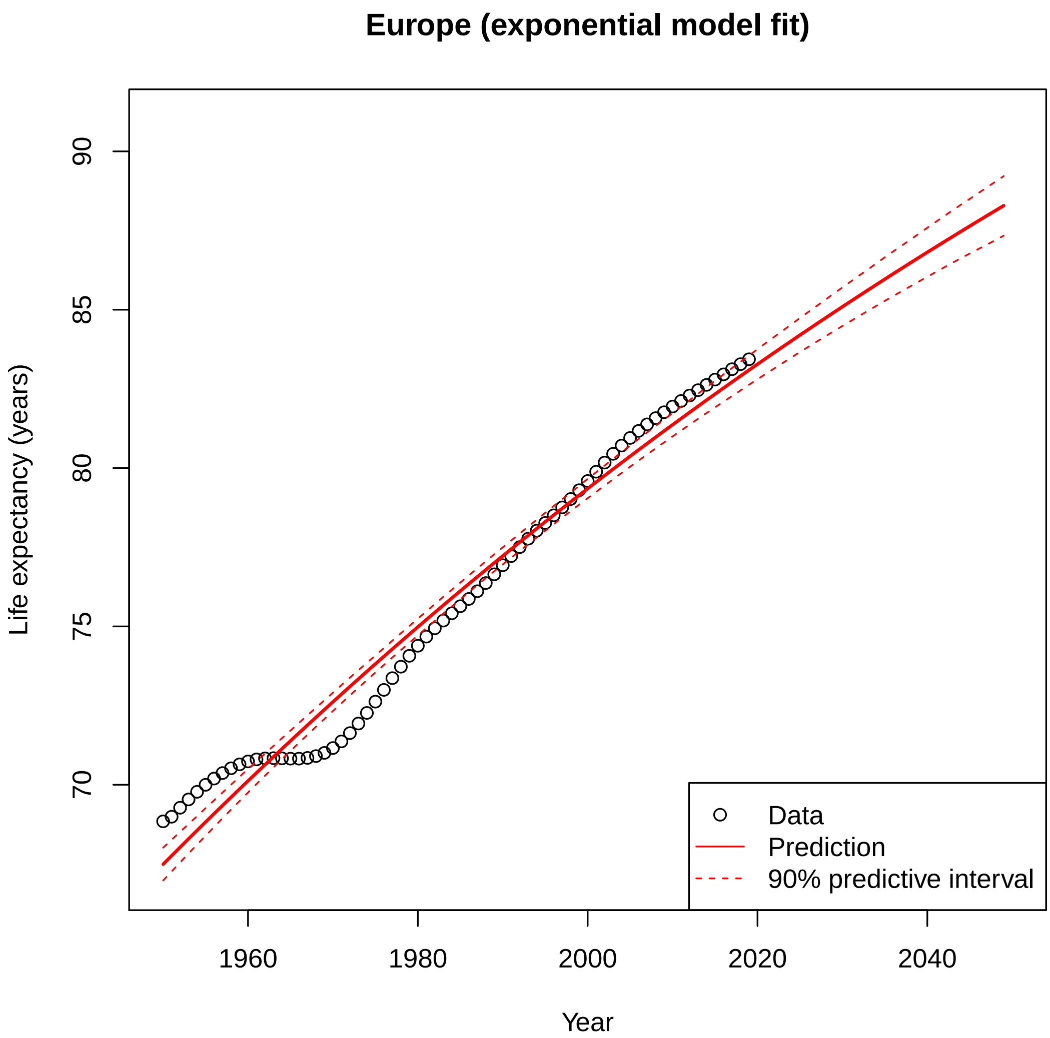

Figure 4: Data, predicted model means, and 90% predictive intervals of life expectancy in Europe between years 1950-2050 using the exponential model (Eq. 7).

4.3 Convergence diagnostics

Rhat is a convergence diagnostic that measures how well the model parameters have mixed. When the Rhat value is less than 1.05 the sample is usable and the closer to 1 the value is the better the convergence in given sample. The Rhat values for these models were taken directly from the stan fit using command "summary(Name of the fit)". The Rhat values for every parameter of every model were 1.00 when rounded off to two decimal places. As that is very close to 1 it is clear that all the chains have converged very well and all models seem to reliable based on Rhat value.

Divergence means that the simulated Hamiltonian trajectory departs from the true trajectory. This means in practice that if there is many iterations that end with a divergence the simulation is biased. Maximum treedepth on the other describes the cap on the depth of the trees that it evaluates during each iteration. Thus exceeding maximum treedepth means that simulation is terminated prematurely and there may be some efficiency problems. Divergences and tree depth were checked using the check_hmc_diagnostics function separately for all models. The function gives the number of iterations that end with a divergence and the number of iterations that exceed the maximum treedepth. It is worth noting that the maximum treedepth was previously set to 15. For the linear model 0 of 4000 iterations neither end with a divergence nor reached the maximum treedepth. This would indicate that the linear model could be trustworthy and is fairly efficient. For the cubic model 0 of 4000 iterations ends with a divergence, but 98 iterations reached the maximum treedepth. The divergence diagnostic is good for the cubic model and only 2.45% of the iterations reached maximum which means that there might be some problem with efficiency but nothing too big. Lastly for the exponential model 0 of 4000 iterations end with a divergence and also 0 iterations reach the maximum treedepth. Thus the exponential model seems also trustworthy and quite efficient based on these diagnostics.

Efficient sample size tells the number of direct samples that would give an estimation of same quality than the model. Low effective sample size means that there are values that dominate the results thus suggesting there might be estimation error. The efficient sample sizes were taken directly from the stan fit as the n_eff value. For the linear fit all of the n_eff values are above 3000 which is very high value considering that the stan fit used 4000 iterations. Thus this indicates that there are no dominating values for any parameters of the linear fit. For the cubic fit the for around half of the parameters n_eff value is between 2000 and 3000 and for the rest above 3000. These values are significantly lower than those of the linear fit, but still overall very high values and there is no reason to believe that there are few dominating values. Finally for the exponential fit n_eff values for the model parameters a, b and λ were the smallest with several values between 1300 and 2000. For the rest of the parameters the n_eff values were mostly over 3000. The exponential fit had the lowest n_eff values, but still all values above 1000 can be considered high and adding to that a good portion of values were over 3000, suggesting that the results are quite reliable and there are not few dominating values.

4.4 Posterior predictive checks

Posterior predictive checks were done using "ppc_dens_overlay" function for all fits. The function overlays the densities of empirical data and the simulated densities of posterior predictive distribution. If the densities of empirical data and simulated data match, it suggests that the models are describing the data correctly.

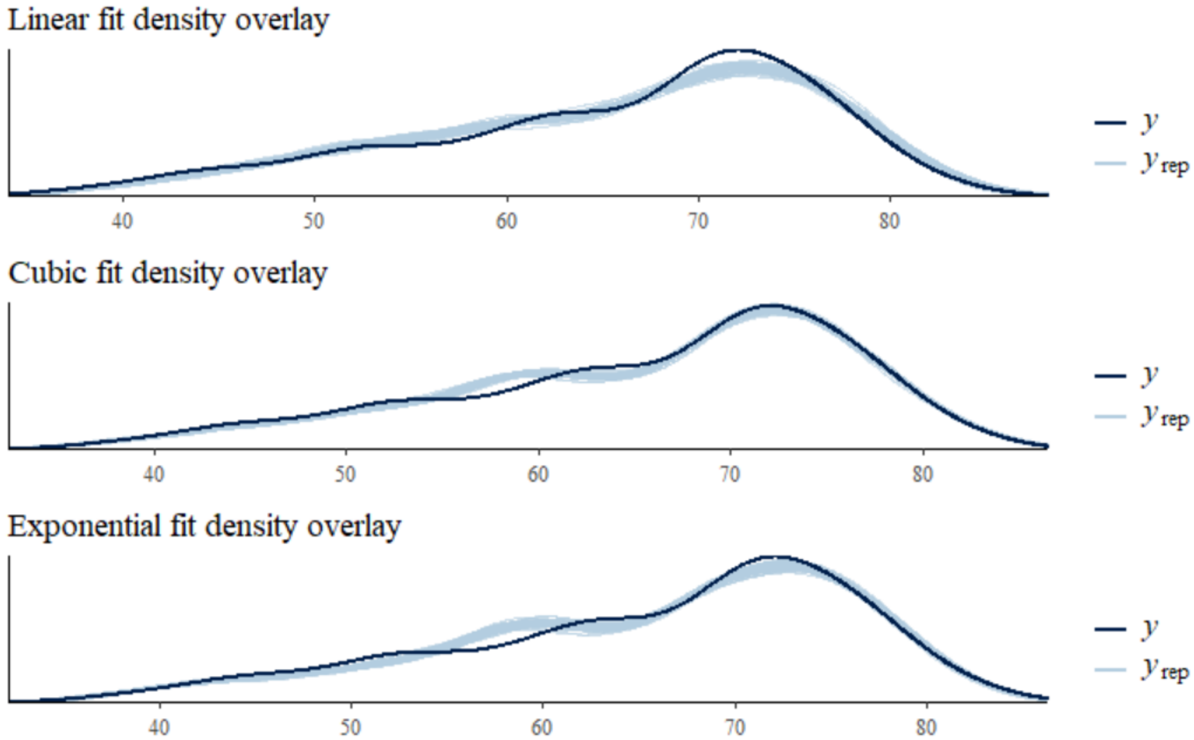

Figure 5: Overlaid density of all life expectancy data described as y and the densities of posterior predictive distributions as y_{rep} for all three fits.

From figure 5 we can see that the posterior predictive densities match the life expectancy data fairly well. For the linear model the only part where the densities don't match too well is 70 to 73 life expectancy which is very small part of the entire data, so linear fit seems to fit the data well. The cubic fit posterior densities differ from the data at both sides of 60, however overall the densities are very similar. Thus it is evident that cubic fit also fits the data very well. Lastly the exponential fit posterior densities differ from the data also at both sides of 60, but a little bit more than the cubic fit posterior densities. There is also minor difference at the 70 to 73 range but that difference is small. The exponential fit posterior density differs a bit more from the data density than the other two models, but still it matches the data density well overall.

4.5 Predictive performance assessment and model comparison

The predictive performance assessment was done by using the loo package in R. The elpd_loo value that the loo function calculates is the Bayesian leave-one-out estimate of the expected log pointwise predictive density and is a sum of individual pointwise log predictive densities. Therefore the larger the value the better the prediction accuracy. For the linear model the elpd_loo value is -1197.3 with a standard error of 32.7. For the cubic model the elpd_loo value is -949.9 with a standard error of 45.6. Lastly for the exponential model the elpd_loo value is -1030.7 with a standard error of 39.5. Thus if these models are compared together the cubic model has clearly the largest value which suggests that it would have the best predictive accuracy before exponential model and linear model. The exponential model should have the second best predictive accuracy and linear the worst of these three. However, when looking at figure 3 we can see that the 90% predictive interval is very wide in the prediction part and the prediction curve itself doesn't look very probable. Thus even though the elpd_loo values suggest that cubic model could have the best predictive accuracy it doesn't seem so by looking at the results. On the other hand exponential models and linear models predictions seem more reasonable as seen from figures 2 and 4.

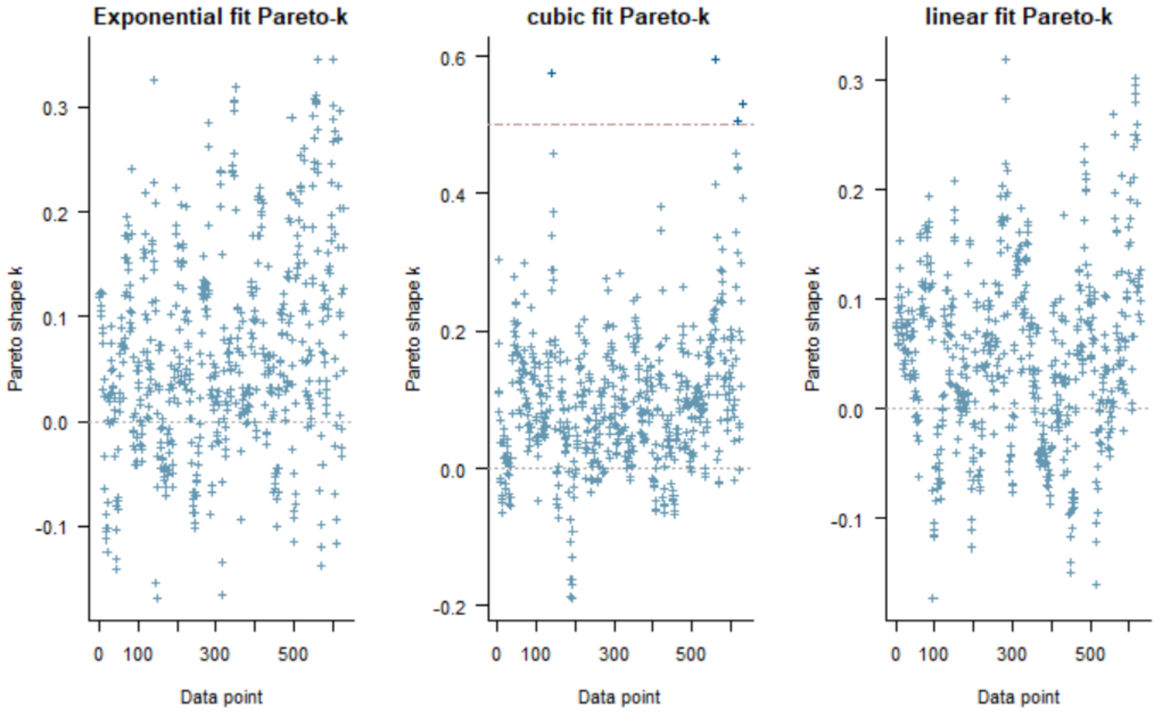

Figure 6: Pareto-k diagnostic values for all three models.

Pareto-k diagnostic measures the reliability of loo function estimates. The lower the pareto-k values are the better and values lower than 0.7 are fine. As seen from figure 6 pareto-k values for linear and exponential models are all less than 0.4 and for cubic model few values were between 0.4 and 0.6. This suggests that elpd_loo values are reliable.

4.6 Prior sensitivity analysis

As the exponential model takes considerable amount of time to run, we were unable to analyse the model sensitivity to all priors separately and instead run the exponential model once with the following tighter priors.

$$
\begin{aligned}
\mu_a &\sim \text{Normal}(0, 20) \\
\mu_b &\sim \text{Normal}(0, 20) \\
\sigma_a &\sim \text{Inv-}\chi^2(1) \\
\sigma_b &\sim \text{Inv-}\chi^2(1) \\
\sigma &\sim \text{Inv-}\chi^2(1) \\
\alpha &\sim \text{Exp}(1) \\
\beta &\sim \text{Exp}(1)
\end{aligned} \tag{9}
$$

The amount of deviation in the prior parameters was chosen quite arbitrarily such that the standard deviations were $5 - 10$ times smaller than in the first model. For comparing the results we can compare some some central parameters posterior distributions with different priors. We can see graphically from Fig. 7 that the priors have some effect on small quantiles, but overall the distributions look very similar.

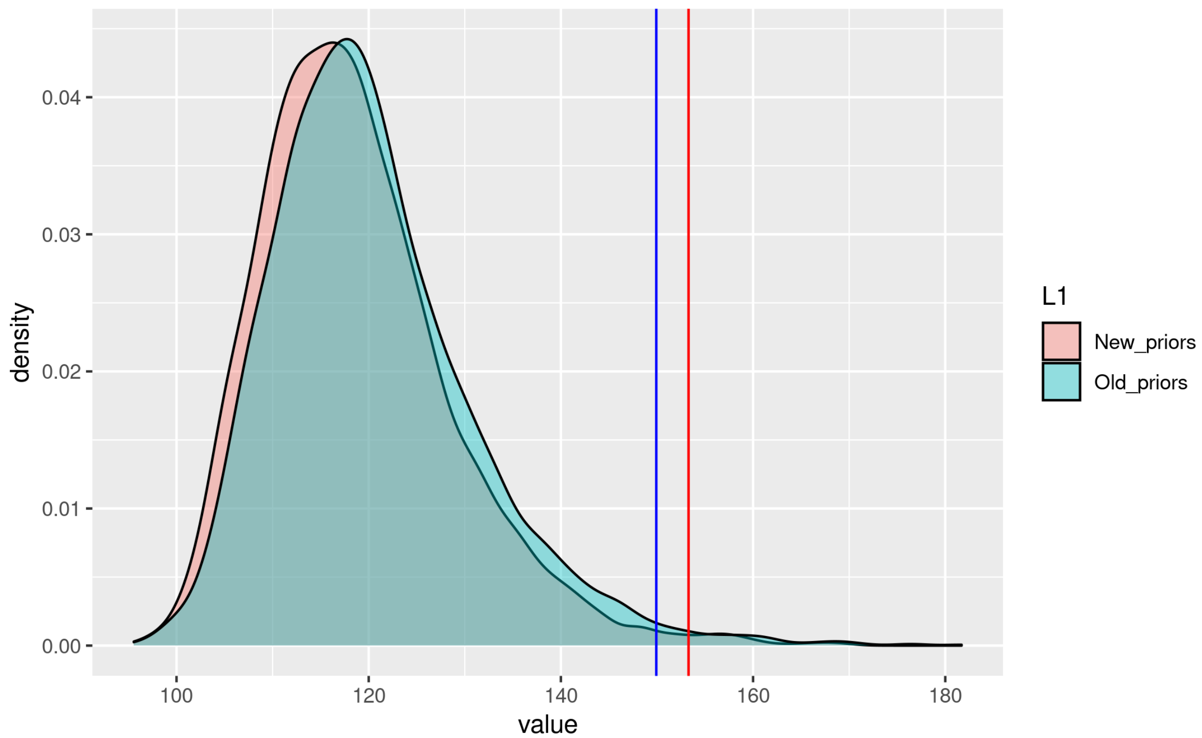

Figure 7: Posterior distributions for a_4 (Europe) parameter in the exponential model with 2 different choices of priors. Also 99% quantiles shown as vertical lines. This is effectively the upper limit of life expectancy according to the fitted model.

PSIS-LOO cross-validation shows little to no difference between the priors as the values are within the standard errors. The only notable difference found was that with the tighter priors, the Pareto k-values were a bit higher, some of them being between 0.5 and 0.7, while with the less informative priors all of the values were under 0.5.

5 Discussion and conclusions

We did have a few minor issues in our project, some of them was the data itself. Our data was little bit interesting, because in some cases the life expectancy in our data between 1950-2020 was totally different than in some other sources. Therefore we could find another data set from better source. However, as the main goal of this project was learning to use stan and the related frameworks, we didn't see this as a large problem.

For these kind of data sets we could have used different models, for example square root or logarithmic models. Also, some other type of model such as time-series model could have been more suitable instead of hierarchical regression model for this kind of time series data. However, time series models were not covered in depth on this course and we didn't have enough knowledge to properly use them.

Another factors we could use to improve the models are other predictors aside from the year which could explain why life expectancy is increasing or decreasing which could be for example GDP (BKT in finnish) per capita or some democracy or stability index. But these are also quantities we can't really predict into the future any better than the life expectancy itself so in this project we decided not to use any additional predictors.

As we can see from figure 7, the more narrow priors we choose have a very little effect on our model. Therefore we could have also used more informative priors to our models to get better predictions.

As we can see from plotted results, in figure 2, 3 and 4, all the models seem to predict well in the short term, but not necessarily in the long term. In linear model we can see that we predict that life expectancy will increase forever, but we know the fact from biology that there is limit for human body to function. Therefore linear model is not realistic in the long term. The cubic model is predicting pretty well until it reaches a local maximum and starts decreasing. This is interesting because this means that there will be some sort of worldwide disaster (corona 2.0?), which is ~~not~~ realistic. This could also mean that the human population starts to follow some kind of oscillating growth curve such as Lotka-Volterra model also seen with animal populations due to human overpopulation and lack of basic resources[6]. The idea of exponential model is best. This is because, the life expectancy will increase by time, but the increase will slow down when it gets closer to the limit a_i. However in our model the increment is reaching the limit quite slowly. Still it seems to be most reasonable model in the long term.

6 Self-reflection

We tried to divide the work in some chunks between team members, but in the end we all learned a lot about both using rstan as well as visualising results with R. Also in this project we had to use almost everything learned during this course in practice which improved our overall bayesian data analysis skills significantly. Of course it was also interesting to learn about the topic of our project itself, how life expectancy has changed in the last 70 years in different areas of the world.

From the stan programming we learned that using any non-standard models like the cubic model is highly inefficient as it took multiple hours to fit on aalto.jupyterhub servers. For comparison fitting the linear and exponential models took somewhere around 5 minutes, though this also varied vastly between runs and chosen priors.

Additionally it was also good practise to read and explain in detail the different convergence diagnostics and predictive checks used in stan as they were one of the most difficult topics on this course.

A Stan code

A.1 Linear model

```
data {
 int < lower =0 > N; // number of data points
 int < lower =0 > Nc; //number of countries/continents
 matrix [Nc,N] y; // observation life expectancies, Nc countries
 int< lower =0 > Npred ; // predict this many years forward
}
transformed data{
  vector[N] x1; // just use years 1,2,...N instead giving x as argument
  for ( i in 1: N){
    x1[i] = i;
  }

}
parameters {
  real alpha[Nc];
  real beta[Nc];
  real < lower =0 > sigma;
  real bmu0;
  real < lower =0 > bsigma0;
  real amu0;
  real < lower =0 > asigma0;
}

 model {
    bmu0 ~ normal(0,10); // yearly change/slope
    amu0 ~ normal(50,20); // we have observations at x = 1 so lets just use something
    bsigma0 ~ inv_chi_square(0.01); // deviation around 10
    asigma0 ~ inv_chi_square(0.01); // deviation around 10
    sigma ~ inv_chi_square(0.01); // deviation around 10 years
   beta ~ normal(bmu0,bsigma0);
   alpha ~ normal(amu0,asigma0);

   for ( j in 1: Nc ){
      y[j, ] ~ normal (alpha[j] + beta[j]*x1 , sigma);
   }
 }

 generated quantities {
    matrix [Nc,Npred] ypred;
```

```
    vector [Nc*N] log_like;
    vector [Nc*N] y_rep;

    for ( j in 1: Nc ){
      for(i in (N+1):(N+Npred)){
        ypred[j,i - N] = normal_rng(alpha[j] + beta[j] * i , sigma);
      }
    }

    for ( j in 1: Nc ){
      for ( i in 1: N){
        log_like[(j-1)*N + i] = normal_lpdf(y[j, i] | (alpha[j] + beta[j] * i) , sigm
        y_rep[(j-1)*N + i] = normal_rng(alpha[j] + beta[j] * i , sigma);
      }
    }
}
```

A.2 Cubic model

```
data {
 int < lower =0 > N; // number of data points
 int < lower =0 > Nc; //number of countries/continents
 matrix [Nc,N] y; // observation life expectancies, Nc countries
 int< lower =0 > Npred ; // prediction year
}
transformed data{
  vector[N] x1;
  vector[N] x2;
  vector[N] x3;// just use years 1,2,...N instead giving x as argument
  for ( i in 1: N){
    x1[i] = i;
    x2[i] = i*i;
    x3[i] = i*i*i;
  }

}
parameters {
  real A[Nc];
  real B[Nc];
  real C[Nc];
  real D[Nc];
  real < lower =0 > sigma;
  real Amu0;
  real < lower =0 > Asigma0;
  real Bmu0;
  real < lower =0 > Bsigma0;
  real Cmu0;
  real < lower =0 > Csigma0;
  real Dmu0;
  real < lower =0 > Dsigma0;
}

 model {
    Amu0 ~ normal(0,0.1);
    Bmu0 ~ normal(0,1);
    Cmu0 ~ normal(0,10);
    Dmu0 ~ normal(0,100);
    Asigma0 ~ inv_chi_square(10);
    Bsigma0 ~ inv_chi_square(1);
    Csigma0 ~ inv_chi_square(0.1);
```

```
    Dsigma0 ~ inv_chi_square(0.01);
    sigma ~ inv_chi_square(0.001);

    A ~ normal(Amu0,Asigma0);
    B ~ normal(Bmu0,Bsigma0);
    C ~ normal(Cmu0,Csigma0);
    D ~ normal(Dmu0,Dsigma0);

    for ( j in 1: Nc ){
      y[j, ] ~ normal (A[j]*x3 + B[j]*x2 + C[j]*x1 + D[j] , sigma);
    }
 }

 generated quantities {
    matrix [Nc,Npred] ypred;
    vector [Nc*N] log_like;
    vector [Nc*N] y_rep;

    for ( j in 1: Nc ){
      for(i in (N+1):(N+Npred)){
        ypred[j,i - N] = normal_rng(A[j]*i*i*i + B[j]*i*i + C[j]*i + D[j] , sigma);
      }
    }

    for ( j in 1: Nc ){
      for ( i in 1: N){
        log_like[(j-1)*N + i] = normal_lpdf(y[j, i] | (A[j]*i*i*i + B[j]*i*i + C[j]*i
        y_rep[(j-1)*N + i] = normal_rng(A[j]*i*i*i + B[j]*i*i + C[j]*i + D[j] , sigma
      }
    }
}
```

A.3 Exponential model

```
data {
 int < lower =0 > N; // number of data points
 int < lower =0 > Nc; //number of countries/continents
 matrix [Nc,N] y; // observation life expectancies, Nc countries
 int< lower =0 > Npred ; // predict this many years forward
}
transformed data{
  vector[N] x1; // just use years 1,2,...N instead giving x as argument
  for ( i in 1: N){
    x1[i] = i;
  }

}
parameters {
  real alpha[Nc];
  real beta[Nc];
  real<lower=0,upper= 1> lambda[Nc];

  real < lower =0 > sigma;
  real bmu0;
  real < lower =0 > bsigma0;
  real amu0;
  real < lower =0 > asigma0;
  real < lower =0 > beta1;
  real < lower =0 > beta2;
}

 model {
    bmu0 ~ normal(0,100); // yearly change/slope
    amu0 ~ normal(0,100); // we have observations at x = 1 so lets just use something
    bsigma0 ~ inv_chi_square(0.01); // deviation around 10
    asigma0 ~ inv_chi_square(0.01); // deviation around 10
    sigma ~ inv_chi_square(0.01); // deviation around 10
    beta1 ~ exponential(0.1); // mean 10
    beta2 ~ exponential(0.1); // mean 10
    lambda ~ beta(beta1, beta2);

   beta ~ normal(bmu0,bsigma0);
   alpha ~ normal(amu0,asigma0);

   for ( j in 1: Nc ){
```

```
      for ( i in 1: N ){
        y[j, i] ~ normal(alpha[j] - beta[j] * pow(lambda[j], x1[i]) , sigma);
      }
    }
  }

  generated quantities {
     matrix [Nc,Npred] ypred;
     vector [Nc*N] log_like;
     vector [Nc*N] y_rep;

     for ( j in 1: Nc ){
       for(i in (N+1):(N+Npred)){
         ypred[j,i - N] = normal_rng(alpha[j] - beta[j] * pow(lambda[j], i) , sigma);
       }
     }

     for ( j in 1: Nc ){
       for ( i in 1: N){
         log_like[(j-1)*N + i] = normal_lpdf(y[j, i] | (alpha[j] - beta[j] * pow(lambd
         y_rep[(j-1)*N + i] = normal_rng(alpha[j] - beta[j] * pow(lambda[j], i), sigma
       }
     }
}
```

References

[1] Vehtari A. et. al. Bayesian Data Analysis Third edition 2021

[2] https://www.statista.com/chart/18826/number-of-hundred-year-olds-centenarians-worl

[3] https://www.kaggle.com/albeffe/life-expectancy-per-country-from-1543-to-2019

[4] https://population.un.org/wpp/Graphs/900

[5] https://population.un.org/wpp/Publications/Files/WPP2019_Methodology.pdf

[6] https://en.wikipedia.org/wiki/Lotka%E2%80%93Volterra_equations