Contents

Abstract **1**

1 Introduction **1**

1.1 Problems Targeted . . . 1

1.2 Methods Involved . . . 1

2 Models and Inference **1**

2.1 Problems Reviewed . . . 1

2.2 Methodology . . . 1

2.2.1 Functional ANOVA . . . 2

2.2.2 Smoothing and Functional response model . . . 2

2.2.3 Functional Time Series Model . . . 2

2.2.4 Funtion-on-function Regression Model . . . 3

2.2.5 Gaussian Process Regression (GPR) Model . . . 3

3 Swedish Lifetable Study **3**

3.1 Description of the Data Set . . . 3

3.2 Exploratory Data Analysis . . . 3

3.3 Scalar-on-function Model . . . 4

3.4 Functional Time Series Model . . . 5

3.5 A More General Function-on-function Model . . . 7

3.6 Model Comparison . . . 9

4 Simulation Study **10**

5 Conclusion **11**

Appendix **12**

Abstract

In this report however, topics from Ch4b, Ch4c, and Ch5 of the STA5004 lectures are included. Other than functional Principal Analysis based method, and scalar-on-function regressions, in this report we are going to mainly focus on the functional linear models including Functional ANOVA model, functional time series model and function-on-function regression model. Comparisons are made between their fitted and prediction performance.

For the Swedish life table data, firstly generalized cross validation (GCV) is applied to find the optimal fitting of the basis function. For different birth years, the logarithmic hazard rate has an evident variation trend with ages. Then, a function-on-scalar regression model is applied to discover the relationship of hazard rate to years. Then, functional time series model and function-on-function are adapted for considering adjacent hazard curve's impact. Both of them provide better predictions than just the mean hazard curve. Comparisons are made through several statistic, ets model and function-on-function model are found to be the optimal under this setting.

Finally, for the simulation section, Gaussian process regression (GPR) model is applied with hyper parameters given, predictions are also made as required.

Keywords: functional time series, function-on-scaler, function-on-function, Gaussian Process Regression

1. Introduction

After exploratory analysis of functional data for fruit flies' egg count has been conducted, this report focus on carrying out several functional regressions in exploring the relationship between functional response and scalar or functional predictors. Functional ANOVA is applied first, then several models are further applied to optimize the fitting result. A simulation data set with application of Gaussian process regression model is shown separately.

1.1 Problems Targeted

This report mainly focuses on three problems: function-on-scalar model, function-on-function model and Gaussian Process model. Problems targeted in this report are:

(a) Smooth these data and find if there is a clear evidence in how they change over time.

(b) Create a functional linear model to predict the hazard curves from birth year. Choose smoothing parameters GCV. Provide a plot of the error covariance. Plot the coefficient functions along with confidence intervals.

(c) Find any indications of lack of fit from residuals. Plot the R-squared for comparing different models. Find any evidence for the effect of time on hazard curves.

(d) Predict the hazard rate at 1920. And evaluate the prediction result with mean hazard curve.

(e) Fit GPR model to simulate data and predict.

1.2 Methods Involved

Methods introduced in order to reach the desired conclusions are provided here. To find the optimal smoothing parameter λ: generalized cross validation (GCV). Summary statistics: the mean curve, covariance matrix, mean square error (MSE), root mean square error (RMSE). R-squared, and etc. Concurrent function-on-scalar regression, function-on-function model and several functional time series model, Gaussian Process Regression (GPR).

2. Models and Inference

After smoothing is done through GCV, this report focuses on finding of scalar or functional predictors' relationships to a functional response, topics of functional linear models, as well as confidence interval and appropriate statistics are used in the first part. In the second part, the GPR model is discussed.

2.1 Problems Reviewed

Other than considering a scalar response, the response variable is functional. The generalized cross-validation (GCV) was adopted for determining the optimal smoothing parameter (λ). Some summary statistics are calculated. The crucial process is to determine the relationship between response and predictors. functional ANOVA model and varieties of functional time series model, function-on-function model, GPR model are applied.

2.2 Methodology

Several methods with respect to a function-on-scalar, function-on-function, functional time series and Gaussian Process Regression are reviewed at first. Statistical inferences respect to each method are also included.

2.2.1 Functional ANOVA

The function ANOVA Model is defined as

$$x_{ij}(t) = \mu(t) + \alpha_j(t) + \epsilon_{ij}(t)$$

Suppose we have a scalar predictor z_i, then the functional model is

$$y_i(t) = \beta_0(t) + \beta_1(t) z_i + \epsilon_i(t)$$

Assume that $\mathrm{E}\epsilon_i(t) = 0$, and $\beta_j(t) = \Phi_j(t) c_j$, and the new least-squares criterion is:

$$\mathrm{SSE}(\beta) = \sum_{i=1}^{n} \int (y_i(t) - z_i \beta(t))^2 \, dt$$

The error-covariance of this model is calculated:

$$C(s,t) = \frac{1}{n-p} \sum \epsilon_i(t) \epsilon_i(s)$$

Confidence Interval

If the y_i all share the same measurement times t, we can look at $\epsilon_i = y_i - z_i \beta$,and estimate:

$$\hat{\Sigma} = \frac{1}{n-k} \sum \epsilon_i \epsilon_i^T$$

then the confidence interval would be:

$$\mathrm{var}[\hat{b}] = \text{ c2bMap} \circ \text{y2cMap} \begin{bmatrix} \hat{\Sigma} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\Sigma} \end{bmatrix}$$
$$\text{y2cMap}^T \circ \text{c2bMap}^T$$

Functional R^2

For evaluating the performance of functional linear regression models, we have a functional equivalent R^2 based on the usual statisc:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

2.2.2 Smoothing and Functional response model

In additioon to estimating the model, using the smooth model would enable us to not consider the smoothness of $\beta(t)$ at first, the model is as followed:

$$\mathrm{PENSSE}_\lambda(\beta) = \sum \int (y_i(t) - z_i \beta(t))^2 \, dt + \sum_j \lambda_j \int \left[L_j \beta_j(t)\right]^2 dt$$

To select the amount of smoothing, leave-one-out cross validation (LOOCV) is used, let $\beta_\lambda^{-i}(t)$ be the model estimated without $y_i(t)$.

$$\mathrm{CV}(\lambda) = \sum \int \left(y_i(t) - \mathbf{z}_i \hat{\beta}_\lambda^{-i}(t)\right)^2 dt$$

We choose λ to minimize the equation above.

Confidence Interval after Smoothing

After smoothing, we could present a new confidence interval, by adjusting the c2bMap as followed:

$$\text{c2bMap } = \left[\sum \int \Psi_i(t) \Psi_i(t)^T \, dt + R\right]^{-1} \circ \left[\sum \int \Psi_i(t) \phi_j(t) dt\right]$$

2.2.3 Functional Time Series Model

The regression by time series model first require decomposing the smooth curves via a functional principal component analysis as followed:

$$x_i(t) = \sum_{j=1}^{\infty} f_{ij} \xi_j(t)$$

where $\xi_j(t)$ is the $j-th$ orthonormal eigenfunction of $\sigma(s,t)$ and the empirical covariance function is

$$\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t))$$

where the fPCs $\xi_1(t), \cdots, \xi_p(t)$ could be derived through several functional principal component analysis method.

h-step-ahead-forecast

Use p fPCs to define h-step-ahead model for $x_{n+h}(t)$:

$$\hat{x}_{n+h}(t) = \mu(t) + \sum_{j=1}^{p} \beta_{n+h,j} \xi_j(t) + \epsilon(t)$$

Popular functional time series model

- Exponential smoothing state space model:

$$\begin{aligned} y_s &= h(z_{s-1}) + k(z_{s-1}) \epsilon_s \\ z_s &= f(z_{s-1}) + g(z_{s-1}) \epsilon_s \end{aligned}$$

- ARIMA (p,d,q)

$$\left(1 + \sum_{k=1}^{p} \phi_k L^k\right)(1-L)^d y_s = \left(1 + \sum_{k=1}^{q} \theta_k L^k\right) \epsilon_s$$

with $\mathbf{z_s}$ as a state vector, and L is a lag operator $L^k y_s = y_{s-k}$.

2.2.4 Funtion-on-function Regression Model

To taken in consideration of historical information, consider the model

$$y_i(t) = \int \beta(s,t) y_{i-1}(s) ds + \epsilon_i(t)$$

And use an integrated squared error objective function with the bivariate basis expansion and bivariate roughness penalty.

$$\begin{aligned} \text{SISE} &= \sum \left[\int \left(y_i(t) - \psi(t) B \int \phi^\top(s) x_i(s) ds \right)^2 dt \right] \\ &= \sum \left[\int \left(y_i(t) - \int \phi^\top(s) x_i(s) ds \otimes \psi^\top(t) \operatorname{vec}(B) \right)^2 dt \right] \end{aligned}$$

$$\begin{aligned} P_{\lambda_s,\lambda_t}(\beta(s,t)) = \lambda_1 \int [L_s \beta(s,t)]^2 dsdt \\ + \lambda_2 \int [L_t \beta(s,t)]^2 dsdt \end{aligned}$$

Note that $vec(B)$ vectorizes B column-wise.

Confidence Interval

The confidence interval could be written as:

$$\operatorname{var}(\hat{B}) = \text{c2bmap} \circ \begin{bmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{bmatrix} \circ \text{c2bmap}^T$$

More Models

2.2.5 Gaussian Process Regression (GPR) Model

The GPR model is given by:

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \quad f(\boldsymbol{x_i}) \sim \text{GP}(0, k(\cdot,\cdot))$$

And the set of data is:

$$\mathcal{D} = \left\{ \begin{pmatrix} y_1 \\ \mathbf{x}_1 \end{pmatrix} \begin{pmatrix} y_2 \\ \mathbf{x}_2 \end{pmatrix} \cdots \begin{pmatrix} y_n \\ \mathbf{x}_n \end{pmatrix} \right\}$$

For the model prediction, let $\mathbf{x}^*$ be a new data point, we want to fit $y^* = f(x^*) + \epsilon^*$. The posterior distribution of $f(x^*)$ given the traning data $\mathcal{D}$ is also a Gaussian distribution, with:

$$\begin{aligned} \mathrm{E}(f(\boldsymbol{x}^*) \mid \mathcal{D}) = \boldsymbol{\psi}^\top(\boldsymbol{x}^*) \boldsymbol{\Psi}^{-1} \boldsymbol{y} \\ \operatorname{Var}(f(\boldsymbol{x}^*) \mid \mathcal{D}) = k(\boldsymbol{x}^*, \boldsymbol{x}^*) - \boldsymbol{\psi}^T(\boldsymbol{x}^*) \boldsymbol{\Psi}^{-1} \boldsymbol{\psi}(\boldsymbol{x}^*) \end{aligned}$$

Where Ψ is the covariance matrix of $(y_1, \cdots, y_n)$. Thus, the predictive distribution of y^* is also Gaussian, with mean above and the covariance

$$\hat{\sigma}^{*2} = \operatorname{Var}(f(\boldsymbol{x}^*) \mid \mathcal{D}) + \sigma^2.$$

3. Swedish Lifetable Study

In this section, we are going to see how hazard curves change with birth years and try to predict the hazard rate with the powerful tools mentioned. The aim is to develop a model for the way in which hazard curves have evolved over 143 years.

3.1 Description of the Data Set

The Swedish lifetable data consist of the log hazard rates (instantaneous risk of death) at ages 0 to 80 for Swedish men by birth year from 1757 to 1900 as a 81 by 143 matrix, as well as Swede1920, a vector of length 80 giving the observed log hazard rate for the cohort born in 1920.

The log hazard rate ranges from the log rates of 5.892524 to −1.245595. The largest hazard rate happens at the birth year of 1768 and at the year of birth. While the smallest is at the 14 years old for the birth year of 1900. It is likely that the hazard rate is both related to the birth year and the age.

3.2 Exploratory Data Analysis

In this section, when fitting the hazard curves, Fourier basis and B-spline basis of order 5 and with equally spaced break points are considered. Introduce the harmonic acceleration ($Lx = D^3x + \omega^2 Dx$, with $\omega = 2\pi/81$) as well as second derivative.

Table 1: Choosing optimal smoothing basis through GCV. The optimal basis chosen is Fourier basis with penalty $\lambda = 0.01$ with GCV score equals 0.007779218.

Second Derivative

	Fourier	B-spline
GCV Score	0.009489434	0.01938791
optimal λ	0.1	0.0001

Harmonic Acceleration

	Fourier	B-spline
GCV Score	0.007779218	0.01987022
optimal λ	0.01	0.0001

Through GCV, the optimal way of smoothing through the Fourier basis with the harmonic acceleration penalty and optimal $\lambda = 0.01$. Then, a same separate smooth is created for Swede1920, and is shown in Figure 1.1 with original and fitted lines. In Figure 1.2, it implies a clearly evident features of the initially abruptly decreasing and slowly increasing hazard rate from 0 to 80 years.

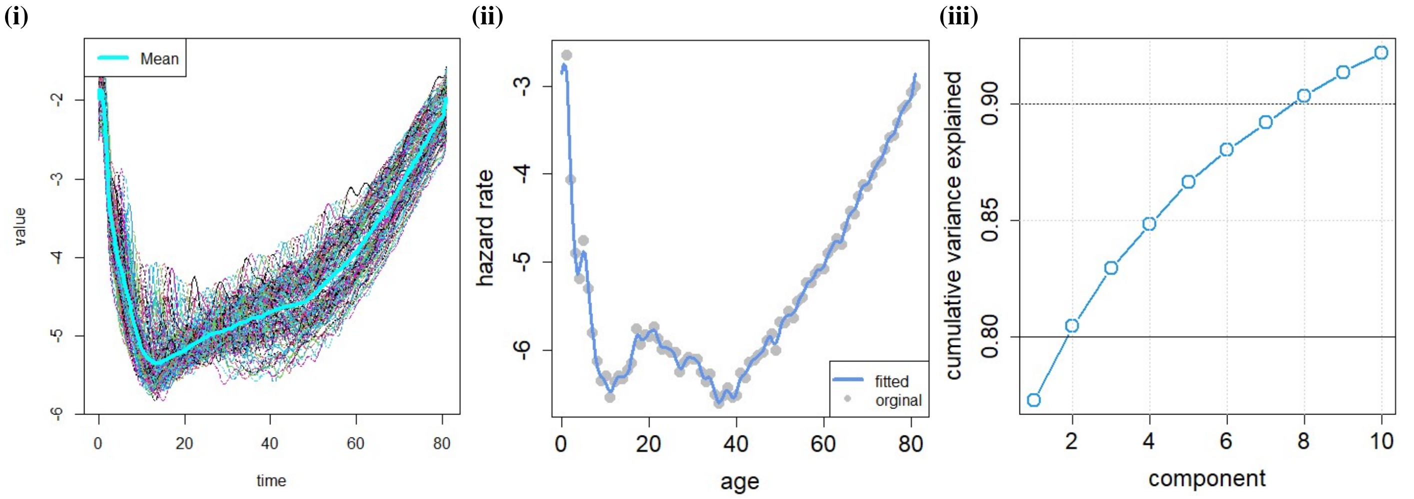


Figure 1: Exploratory analysis of Fourier fitted curves of fruit flies data. **(i)** provides the smoothing of the log hazard rate for years between 1757-1900 given by Fourier basis with second order penalty of $\lambda = 0.01$. **(ii)** provides a same fitting to the birth year of 1920. **(iii)** shows the cumulative plot of fPCA, with lines indicating 80% and 90% of variances explained.

In later sessions, since functional principal component analysis (fPCA) is applied when doing functional time series model, a functional PCA is also conducted a priori for illustrating it's feasibility. In Figure 1.3, while 7 fPCs account for 90% of total variance, 2 fPCs already gained 80% of variance explained. It seems that using 2 fPCs for the functional time series regression would be appropriate.

3.3 Scalar-on-function Model

To figure out the relationship between the hazard rate and birth year, first use a scalar-on-function model.

Functional ANOVA Model

The first attempt that I've taken is using a functional ANOVA Model, but here we only have scalar variables. The model, denoted as $\mathbf{M_1}$ is constructed as $y_i(t) = \beta_0(t)+\beta_1(t)z_i+\epsilon_i(t)$, where z_i represents the nominal variable of years.

In practise, it would be better to introduce the smoothing penalty to the model, avoiding the need to determine the smoothness of βs at first. In Figure 2.1, the LOOCV chooses $\lambda = e^4$, when the minimum cross validation error is 3.15074. The regression coefficients of βs are also shown. Figure A1 is the intercept β_0, it is very similar to variation of the average curve. Figure A2 is β_0, which shows an increase with age at the beginning, then continue to decrease until about 52 years old, then started to increase. While the fitted value of the hazard curves seems ambiguous, for there all of the curves have same trend of variation within different years.

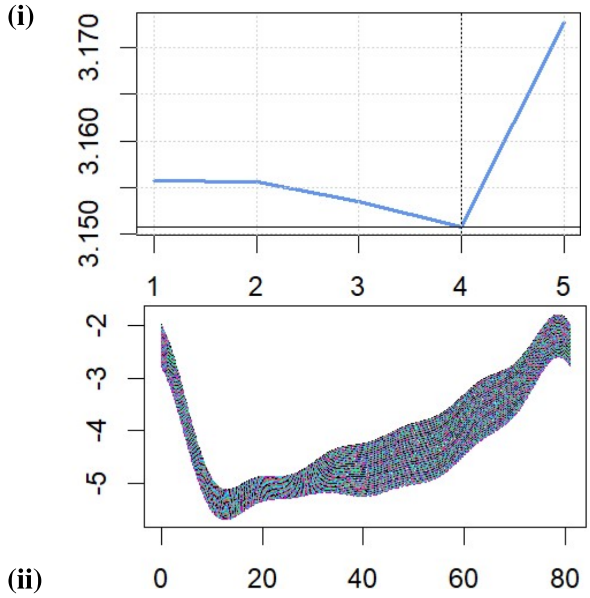


Figure 2: Use LOOCV to find optimal smoothing of functional response model. **(i)** shows optimal $\lambda = e^4$ for model $\mathbf{M_1}$. **(ii)** is the fitted value of hazard rate for birth years 1757-1900, while all the curves have similar trend.

It is possible to derive the confidence interval, since for the highly smoothed curves, the 95% confidence intervals before and after smoothing are very close, only the smoothed ones are shown. For β_0, it has very slim confidence bands, while for β_1, it has a wider band approximately before 20 years old compared to the others. This is also possible to be inferred from Figure 3.1, 3.2.

Then, I carried out some residual diagnostic. The error covariance matrix would provide a glimpse to see if there is any lack of fit in our model.

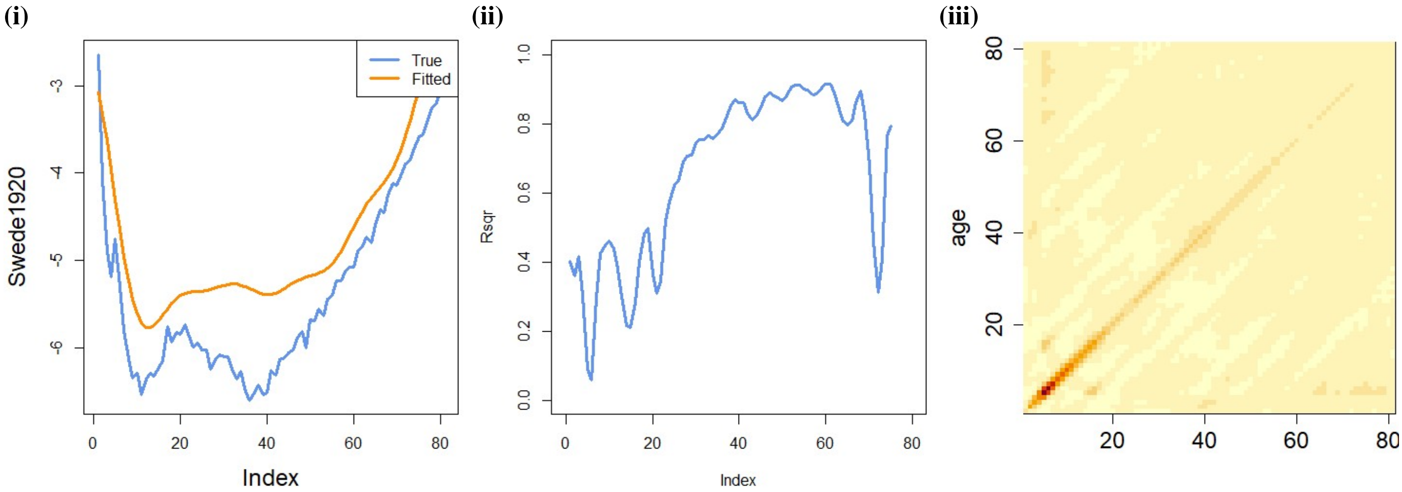


Figure 3: Prediction and Error-Covariance of $\mathbf{M_1}$ model. **(i)** shows the prediction compared with smoothed curve of 1920 cohort. **(ii)** is the R^2 calculated, it shows better fitting for larger ages, and poorly fit with young ages. **(iii)** is the the error-covariance plot, with darker color indicating large variance, it happens on the diagonal. 3-dimensional plot see Figure A6

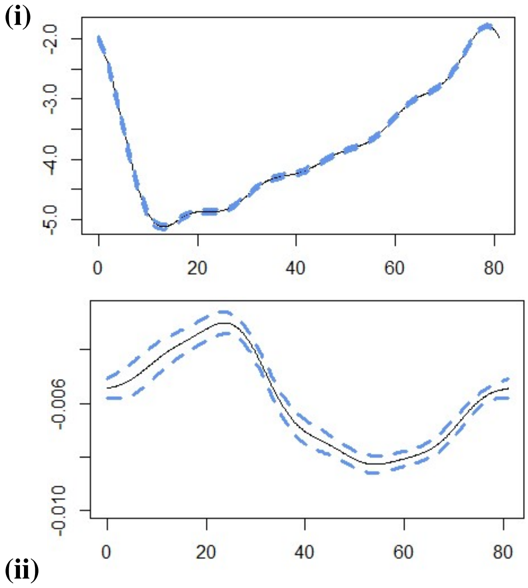


Figure 4: Fitted βs and 95% confidence interval in dashed lines. **(i)** β_0 has very slim confidence interval. **(ii)** β_1 has a wider confidence interval, especially before 20 years old.

The heat map in Figure 3.3 implies a high correlation between the adjacent years. It is very likely that, the population having similar age at a specific year, share similar instantaneous rate of death. This could be further noted in a 3-dimensional plot as Figure 4, A3. It shows the ridge at the diagonal and plateu at two sides.

Further, we use $\mathbf{M_1}$ to predict the 1920 cohort of hazard rate, and draws the R^2 plot, which indicate a relatively good fit with ages between 40 to 60 oscillating above 0.8, and poor fit around boundaries. By using the Kolmogorov-Smirnov test ($D = 0.50617$, $p -$ value < 0.001), two curves have different distribution. Thus, based on both the error-covariance analysis and test result, a new model is proposed to refine the result.

3.4 Functional Time Series Model

The regression model $\mathbf{M_1}$ is ambiguous, for it has similar predicted trend for different birth years. Then, a functional input and functional out-put regression is considered from $x(t)$ to $y(t)$. In the functional time series model, based on the previous result of adequacy of using only 2 fPCs, in this section, only two fPCs are selected for each model.

Basic idea of this process is to firstly decompose the smooth curves via a fPCA, then fit a univariate time series model to fPCs. Forecast the PC scores using the fitted time series models. Then, multiply the forecast PC scores by fixed PCs to obtain forecasts.

In Figure 5, it presents the first two functional principal components and their associated principal component scores. The bottom panel of Figure 1 also plots the forecasted principal component scores, and the 80% prediction intervals (in yellow color), using an exponential smoothing state-space model. The prediction intervals widen very quickly. This reflects the difficulty for the model in forecasting medium or long term horizon, as a result of the increase in variability.

The main effects are very similar to the average curve. And the basis function of first two fPCs are

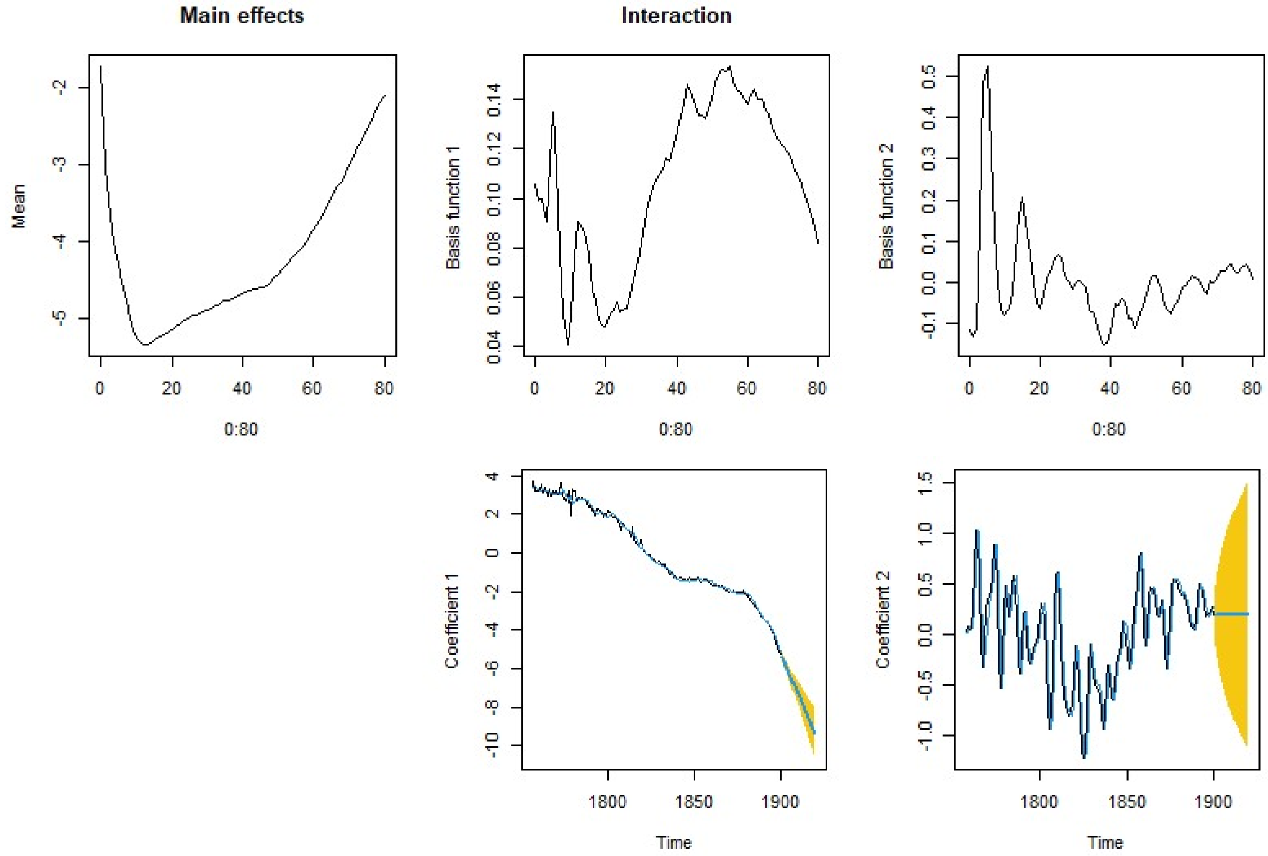


Figure 5: The first two functional principal components and their associated principal component scores for the swedish hazard rate data from 1757 to 1900.

shown. By using an exponential smoothing method, the principal component scores are forecasted. Coefficient of each fPCs are shown, as well as a prediction interval estimated through smoothing errors and model residual error. For the first fPC, increasing years have negative impact on the hazard rate.

There are two ways of making predictions for hazard curve of 1920 cohort: through a 20-step-ahead prediction ($\mathbf{M_2}$), or through iterative one-step-ahead prediction model $\mathbf{M_3}$ (Figure A). In Figure 6.1, the rainbow curves are prediction made from 1900-1920 by a 20-ahead prediction, with the mean hazard curve and real 1920 curve plotted. Compared with the mean curve, the prediction is evidently closer. Albeit the method is different, iterative one-step-ahead prediction provide similar result (Table 2), mildly less than the performance of $\mathbf{M_2}$.

For further exploration, other functional time series model: random walk drift ($\mathbf{M_4}$) and ARIMA ($\mathbf{M_5}$),are applied for comparison. Both of the models have very minor difference in fitting, and mild difference in prediction. Further, KS-test suggest these series have very similar distribution ($P-\text{value} > 0.05$). And the prediction is shown in Figure 7.3, with ets model ($\mathbf{M_2}$) having a seemingly closer prediction, which could be further illustrated as in Table 2.

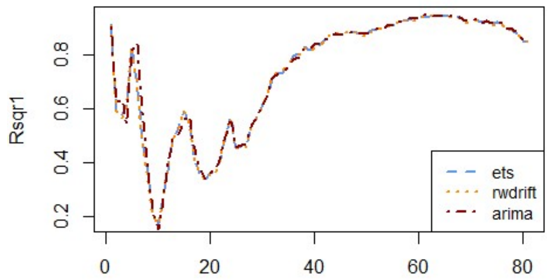


Figure 6: R-square plot of $\mathbf{M_2}$, $\mathbf{M_4}$ and $\mathbf{M_5}$. Both curves overlaps a lot. (for detailed explanation, refer to A5)

As in Table 2, I take an attempt based on the idea from the lecture of data registration, we would be able to make proper measurements of distance between curves with MSE, MAE, RMSE, L^∞, dynamic time warping and etc. Among the five measures, the exponential smoothing state space model ($\mathbf{M_2}$), with 20-ahead forecast has the most optimal prediction ability, with $MSE = 11.30$, $RMSE = 3.36$, $MAE = 24.20$, $L^\infty = 0.85$, $DTW = 15.71$. But it is still interesting that functional ANOVA has a more similar trend to the variation of true 1920 cohort, even if the other measures are large.

Table 2: Functional Time Series Model Comparison.
Traditional statistics such as MSE, RMSE, MAE, L^{∞} and distance with registration of data: dynamic time warping (DTW) are considered for evaluting the distance between 1920 cohort hazard curve. The maximum value of each measurement is highlighted.

Model	MSE	RMSE	MAE	L^{∞}	DTW
ets (h-step, $\mathbf{M_2}$)	11.29514	3.360824	24.19782	0.8514893	15.70854
ets (iterative, $\mathbf{M_3}$)	13.94577	3.734403	25.65752	1.108682	15.87354
rwdrift ($\mathbf{M_4}$)	24.00749	4.899744	37.30823	0.9303991	19.14362
arima ($\mathbf{M_5}$)	14.80427	3.847632	26.34663	0.9086034	15.4658

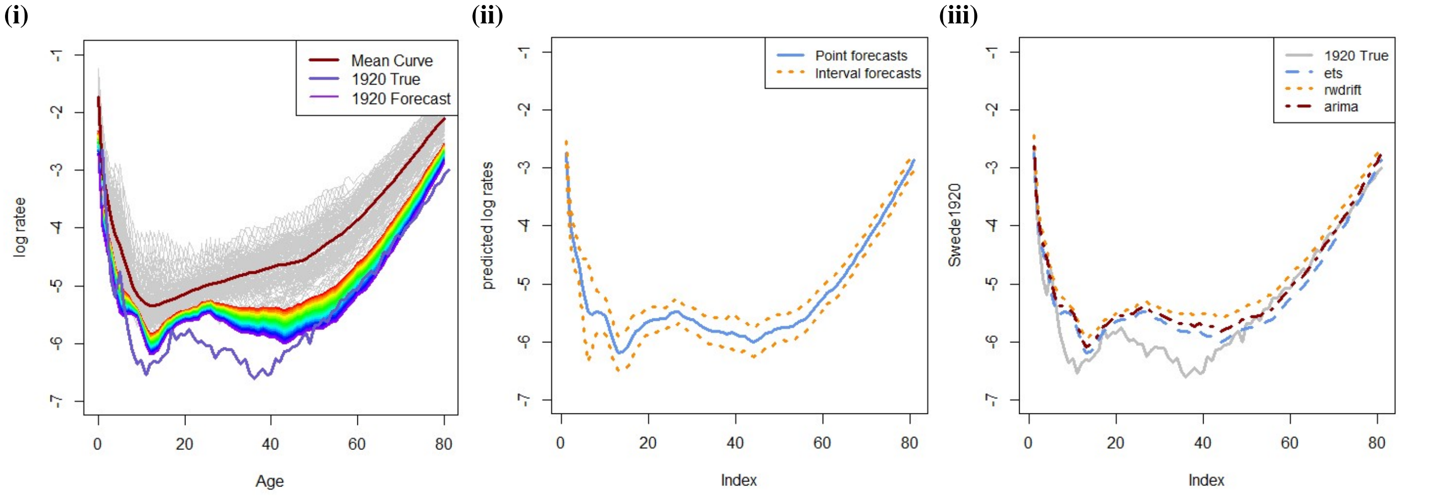


Figure 7: Prediction and Error-Covariance of $\mathbf{M_1}$ model. **(i)** shows the prediction compared with smoothed curve of 1920 cohort. **(ii)** is the R^2 calculated, it shows better fitting for larger ages, and poorly fit with young ages. **(iii)** is the the error-covariance plot, with darker color indicating large variance, it happens on the diagonal.

3.5 A More General Function-on-function Model

The error-covariance matrix further indicates a lack-of-fit of the diagonal phenomenon that population with similar ages, though born in different years might have strong correlations.

Thus, considering combining historical information, a second model $\mathbf{M_6}$ is proposed as: $y_i(t) = \int \beta(s,t)y_{i-1}(s)ds + \epsilon_i(t)$. Here, we seek to use the last year's hazard rate and β containing all of the historical information to do a full integration model of $\beta(s,t)$.

Since $\beta_1(s,t) = \phi'(s)\mathbf{B}\psi(t)$, the way I construct the model is to provide a B-spline basis of order four with 23 basis functions to define functional parameter objects for β_0, $\beta_1(\cdot,t)$ and $\beta_1(s,\cdot)$. The second derivative is penalized in each case, but the smoothing parameter values vary as shown.

In Figure 8, it displays the estimated regression surface $\beta(s;t)$. The estimated intercept function β_0 ranged over values four orders of magnitude smaller than the response functions (Figure 9.2), and can therefore be considered to be essentially zero.

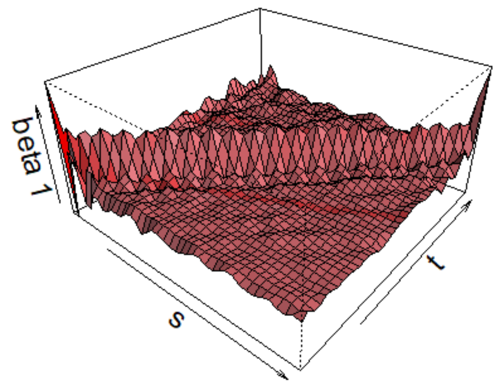


Figure 8: The bivariate regression coefficient function $\beta(s,t)$ estimated from the 143 log hazard rate functions. The ridge in $\beta_1(s;t)$ is one year off the diagonal.

The strong ridge one year off the diagonal, namely $\beta_1(s1;s)$, indicates that mortality at any age is most strongly related to mortality at the previous year for that age less one, which is also illustrated from the countour plot in Figure 9.3. In other words, mortality is most strongly determined by age-specific factors like infectious diseases in infancy,

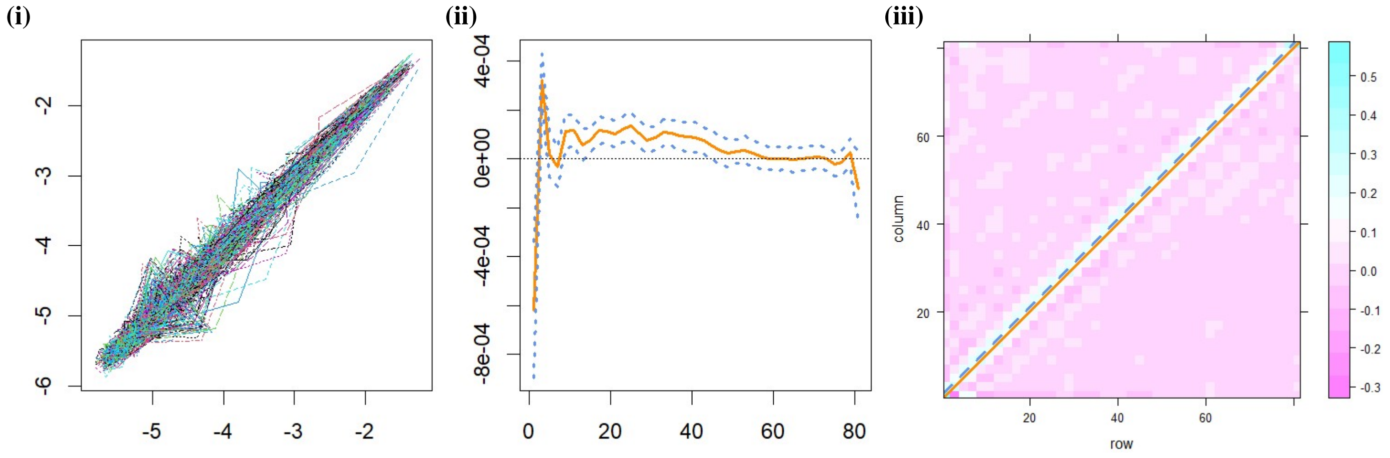


Figure 9: Estimation of M_6 model with bivariate $\beta(s,t)$. **(i)** shows the hazard rate and confidence interval of last year versus next year, The curves are laying near diagonal. **(ii)** shows the varation of β_0, is too small to be considered as zero. **(iii)** shows the levelplot with large β_1 value colored in blue and small value colored in pink. The yellow line is the diagnol, while the blue dashed line crossing through the ridge is one year off the diagonal.

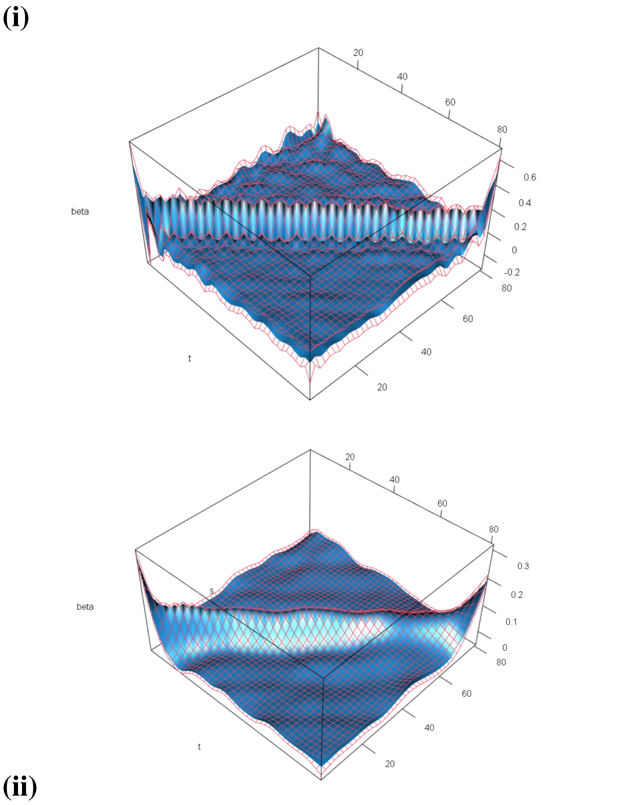


Figure 10: β_1 estimated coefficient surface and confidence interval. **(i)** shows the estimation with $_s = 1$ and $_t = 1$. **(ii)** shows the estimation with $\lambda_s = 1000$ and $\lambda_t = 1000$. Differences in these pots indicate that confidence interval is evidently smaller if we choose larger penalties.

accidents and violent death in early adulthood, and aging late in life. The height of the surface declines to near zero for large differences between s and t for this reason as well.

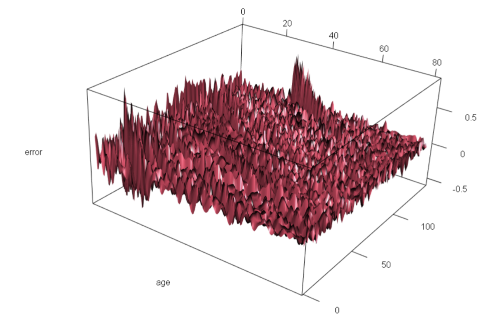


Figure 11: Residual plot with year and age. The chaotic distribution indicates residuals have no linear trends over time.

The confidence interval of β_0 is provided in Figure 9, with a very narrow interval at the beginning and a smooth interval along with mild oscillations of β_0. For β_1, we need to use the Kronecker product basis, where $\Phi(s)\otimes\Psi(t) \longrightarrow x(t,s)$ and rows plotted horizontally, this is shown in Figure 10.1, with red webs as upper and lower interval. Making λ larger can reduce the intervals for other components. E.g., if $\lambda_s = 10^3$, and $\lambda_t = 10^3$, we again plot the confidence interval for β_0 and β_1, they are significantly narrower than previous estimation as in Figure 10.2, A5, the red webs are very close to the surface. For determining the optimal cases, we will need a cross-validation process, but there will be three smoothing penalties.

Next, we also carry out some residual exploration, first we plotted out the residuals and find no clear trends with age from Figure 12.1. If we again look at the error-covariance plot, it is mostly centered very directly on residuals (Figure A7). As for the plot of erros versus years and ages, it shows no clear trend over time as in Figure 11,12.2, there isn't clear trend over time. These results imply that our model has a good fit of the structure of functional data. Then, we would also be possible to find if there is any relationship between residuals and predicted values. If we try model with different shift of 1,11,31,51,71, the model with 1 shift has the most randomly distributed residuals for every prediction value (Figure 13). This indicates that the 1-shift model $\mathbf{M_6}$ used is the optimal.

Fitted performance of evaluated by R^2 is in Figure 14.2, and the predicted curve of 1920 cohort in Figure 14.1. In later sections, we will compare it with models from sections.

3.6 Model Comparison

From 3.3 to 3.5, in each of the section, select the optimal model $\mathbf{M_1}$, $\mathbf{M_2}$ and $\mathbf{M_6}$. A brief summary is provided here for comparing between these three.

For the fitting result, we plot the R-square for each of the three models. And for the prediction result, calculate the MSE, RMSE, MAE, L^∞ distance and (dynamic time warping) DTW distance to all of the 6 models to find their similarity with the Swedish 1920 cohort's logarithmic hazard rate.

Model $\mathbf{M_1}$ has residuals that indicates a lack-of-fit with time. $\mathbf{M_6}$ take in considerations of historical data, thus achieving a better R^2 for most ages. However, functional time series has a better performance in prediction.

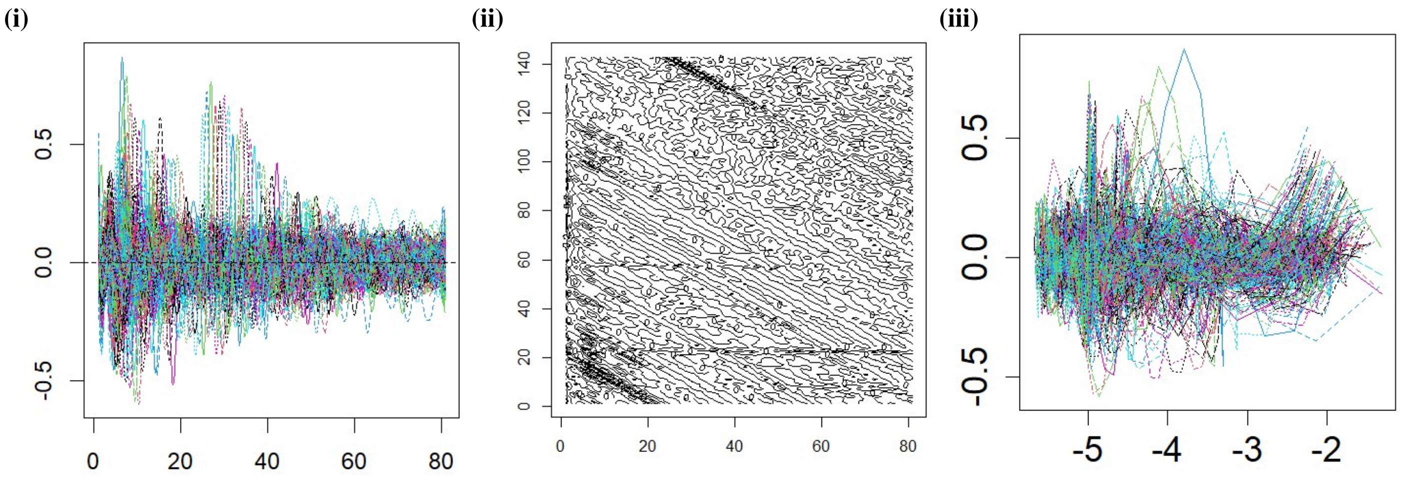


Figure 12: Residual analysis for model $\mathbf{M_6}$. **(i)** shows the residuals with the variation of age, no clear pattern is found. **(ii)** shows the contour plot for residual with age and years, it is chaotic and random, **(iii)** shows the residuals' relationship to predicted values, in this case, we use last year to estimate next year. Both of the residual plots suggest we have achieved a good fit.

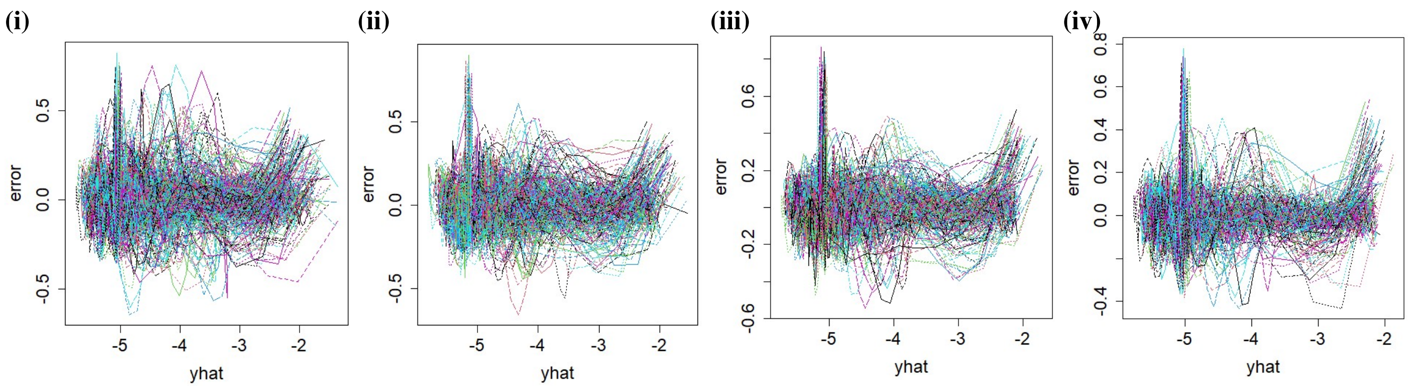


Figure 13: Residual analysis for model $\mathbf{M_6}$ for different number of shift. **(i)** uses 11 shift, **(ii)** uses 31 shift, **(iii)** uses 51 shift, **(iv)** uses 71 shift. From left to right, the residuals first randomly distributed, then has shown a larger residual at the beginning, probably since 71 ago adolescent have not been born. The 1-shift model $\mathbf{M_1}$ is the optimal.

(i)

(ii)

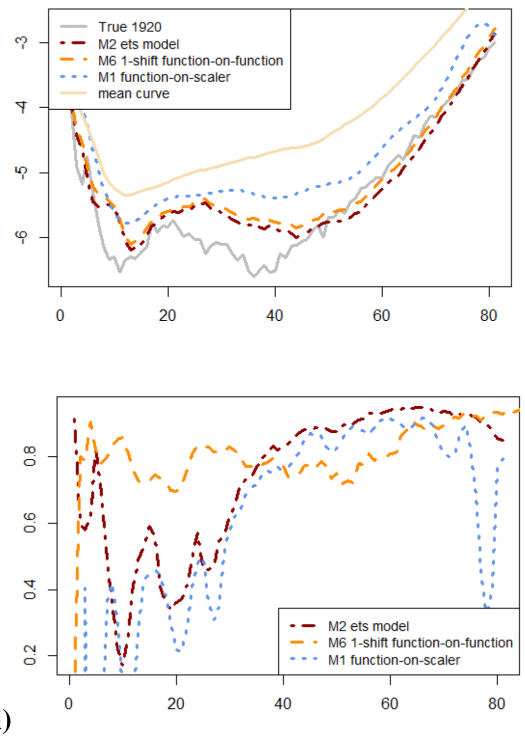


Figure 14: Comparisons of $\mathbf{M_1}$, $\mathbf{M_2}$ and $\mathbf{M_6}$ model. **(i)** shows the predicted 1920 hazard rate, with real value and mean curve plotted. All models are better than mean hazard curve. **(ii)** shows the R^2, M_6 has a very steady and high R^2 compare to others.

All of the model shown has better prediction than the mean hazard rate curve, but some are even better, like $\mathbf{M_2}$ (Figure 14.1). Some has steadily higher R^2, like $\mathbf{M_6}$ (Figure 14.2), especially between the age of 2 to 30 than other models. To consider fitting, we should undoubtedly choose $\mathbf{M_6}$, while for prediction $\mathbf{M_2}$ is better, no matter which kind of measurements that chosen among the five common distance measures. Table 3 provides the result, indicating that $\mathbf{M_2}$ not only close to the real 1920 cohort in distance, but having similar shape of variation compared to other two models.

Table 3: Prediction result evaluation.
$\mathbf{M_1}$ a function-on-scalar model, $\mathbf{M_2}$ ets functional time series model and $\mathbf{M_6}$ -shift function-on-function model. Measurements include MSE. RMSE. MAE, L^∞ and DTW.

	$\mathbf{M_1}$	$\mathbf{M_2}$	$\mathbf{M_6}$
MSE	37.86134	11.29514	20.29558
RMSE	6.153157	3.360824	4.505062
MAE	50.38739	24.19782	33.69434
L^∞	1.281495	0.8514893	0.9186271
DTW	25.51516	15.70854	23.69228

4. Simulation Study

Instead of using the real-world data, this session uses a simulated data to study Gaussian process regression. It involves 4-dimensional covariates. The data is generated as:

$$y = 0.2x_1 \times |x_1|^{1/3} - 4\sin(x_2) + \exp(x_3 \\ + \log(x_4) + \epsilon, \quad \epsilon \sim GP(0, k(\cdot,\cdot))$$

where $k(\cdot,\cdot)$ is specified by the squared exponential covariance function which depends on $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^T$.

(i) (ii) (iii)

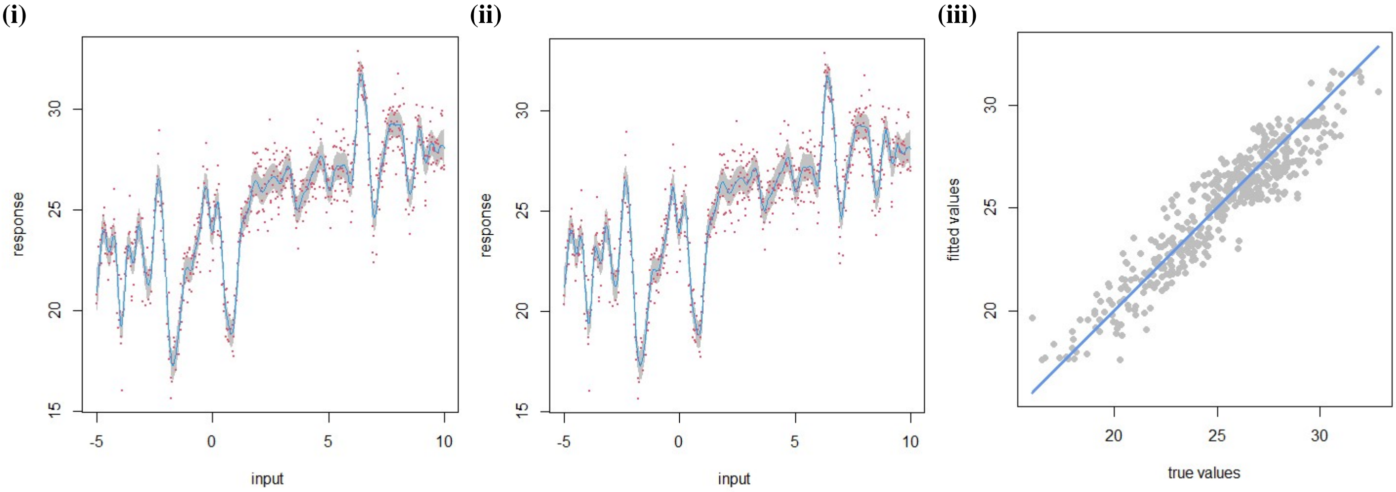


Figure 15: Fitted values and denoising prediction, with 95% predictive interval. **(i)** is the fitted plot. **(ii)** is the denoising prediction curve. **(iii)** is the plot of fitted versus predicted, the line indicating they are very close.

Seed of 1 is set up for reproducing the results. The model training and prediction calculating for the test data are carried out by gpr and gprPredict function from R package GPFDA. Trace of 2 is selected to print out trace step of the optimization process. This process converges from the model output. Among 1000 samples, 360 are chosen to be test set and 640 as training. The estimates of the hyper-parameters are as in Table 4.

Table 4: parameter estimation from GPR
Parameter estimation of the hyper-prior distribution for the hyper-parameters.

Parameter	1^{st}	2^{nd}	$3^{r}d$	4^{th}
linear.a	-0.8809	-4.0101	-8.5615	5.7559
linear.i	6.2346			
pow.ex.v	3.8609			
pow.ex.w	-5.8278	-4.4293	4.6807	1.6818
vv	1.0730			

Figure 15.2 shows the fitted values and the predicted value against one covariate. Fitted curves and predicted curves have very similar variation with input values. Fitted and predicted values are located near to the diagonal as in Figure 15.3. For this model, calculating $R^2 = 0.8505036$, which indicates a rather goof fit.

5. Conclusion

In this report, two main parts are done: the Swedish life table study and simulated data Gaussian process regression study. For a functional response, functional linear models including function-on-scalar, functional time series model and general function-on-function model are considered. For function on scalar model, a functional ANOVA model is considered. But it has ambiguous fitted curves and residual plots, then some function-on-function models are applied. Several functional time series are conducted and compared, the exponential smoothing state space (ETS) model shows the best fitting and prediction.

To solve the lack-of-fit in the residuals, a general function-on-function model of 1 shift is conducted, having better performance than other shifts. The coefficient plot shows a strong relationship between different birth years' adjacent ages. Proper explanations are made with regard to this phenomenon. In summary, one of the functional time series model has a better prediction to the real 1920 cohort's hazard rate, but function-on-function model has the best fitting performance, and relatively well prediction.

In a separate section, a Gaussian process regression model is constructed, parameter estimation and prediction result is provided, it also indicates a convergent and nice fit.

Appendix

Figure A1: the intercept β_0 after LOOCV for $\mathbf{M_1}$.

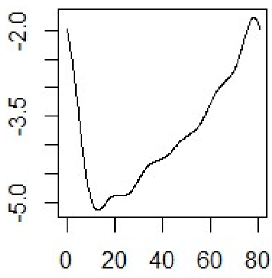
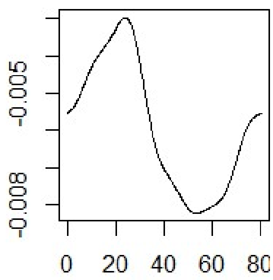


Figure A2: the intercept β_1 after LOOCV for $\mathbf{M_1}$.



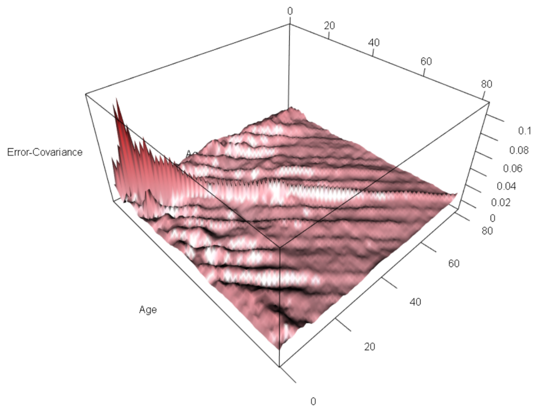


Figure A3: 3-dimensional Error-Covariance plot of $\mathbf{M_1}$ model.

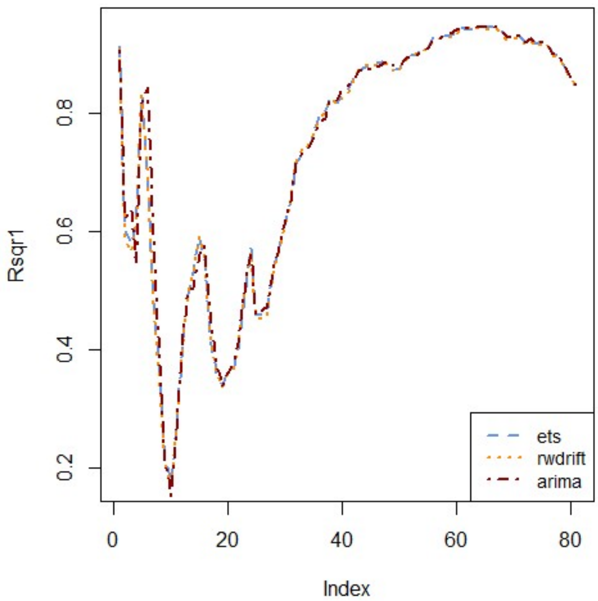


Figure A4: R^2 comparisons between 3 functional time series model.

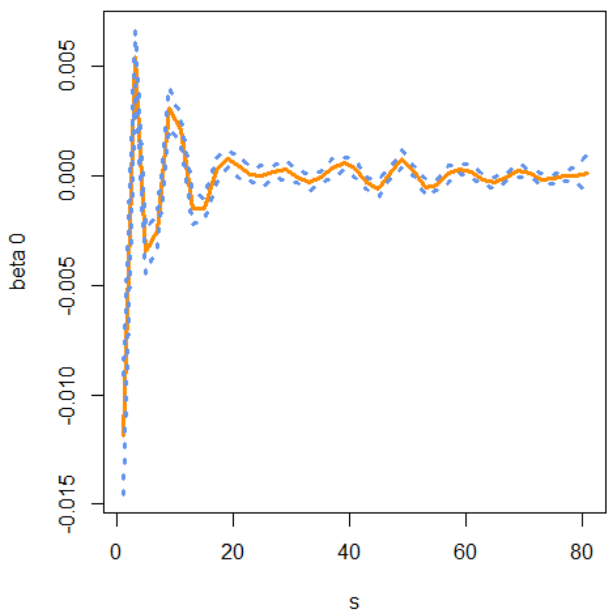


Figure A5: The prediction and Confidence interval for β_0, when ${}_s = 1000$, $\lambda_t = 1000$.

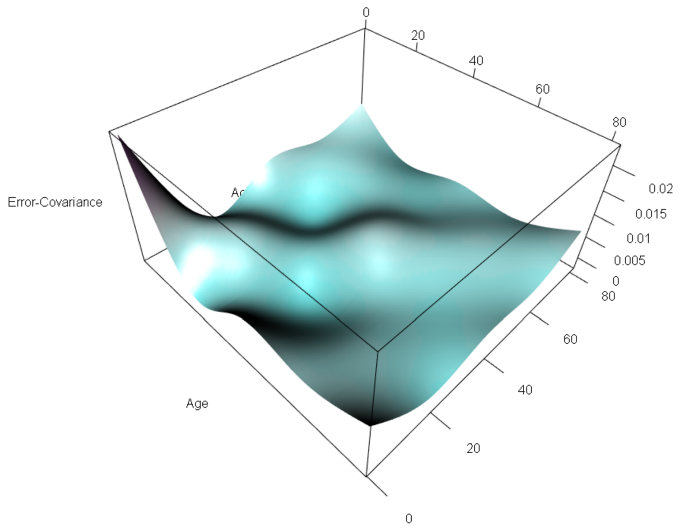


Figure A6: Error covariance plot of $\mathbf{M_1}$ model smoothed by GAM model in R.

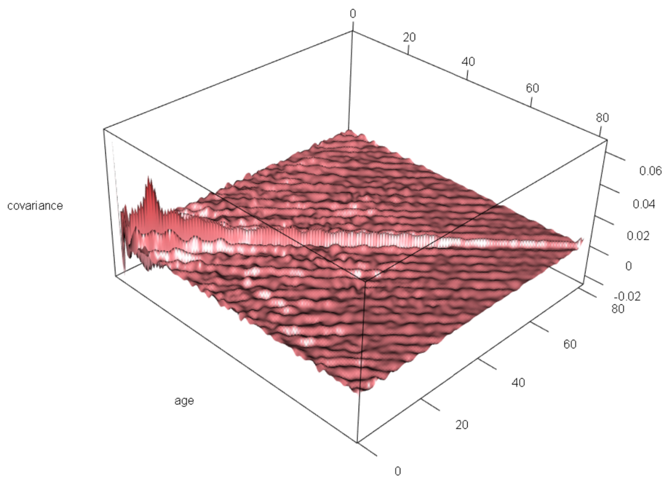


Figure A7: Error covariance plot of model $\mathbf{M_6}$.