

Nonlinear regression analysis of Exponential, Von Bertalanffy and dynamic carrying capacity models for Canada population

Ailiza Aiben

May 4, 2023

Supervisor: Dr.Dongming Wei
Second Reader: Dr.Ugurlu Kerem

Abstract

This project is aimed to analyze the exponential model, Von Bertalanffy model, and the dynamic carrying capacity model based on the Canada population data. The unknown parameters in these models would be found out by using a special function in R language called \sim nls (Nonlinear Least Squares). The mathematical theory of the nls function would be introduced and explained. In order to examine if the models are appropriate for the Canada population, residual plots of the three models will be found out to compare. Furthermore, comparison method AIC (Akaike's Information Criteria) would be used to determine the best model fitted to the Canada population among the three models. The model which shows the best result could be used as an actuarial life contingency model for the Canada population data in the future.

1 Introduction

Exponential model is widely used growth model which should fit to any population data in normal situation. In this project, it will be used as an compared model with the other two models. Because Von Bertalanffy and dynamic carrying capacity models are not used for population data commonly, thus the main aim is to examine these two models based on Canada population data. Von Bertalanffy model was proposed by Von Bertalanffy in 1938 for animal growth as a mechanistic model. This model was extensively used for fisheries research by people. And it is similar in shape with the other widely used growth model called Gompertz model[8]. Gompertz model was used for different human population data by many researchers. Dynamic carrying capacity model is mostly used in the field of ecology by biologists and ecologists. It was firstly introduced by the ecologist C.S. Holling in 1973, which primarily used to observe the animal population growth[5]. Moreover, it is also used in the field of tumor growth such as in the paper of Sebastien in 2014[2]. In this project, the two models will firstly used for Canada population growth. Thus, this paper will contribute to the research of finding models' suitability to different fields.

This project studied three models such as Exponential model, Von Bertalanffy model, and Dynamic carrying capacity model based on Canada population data. In the section 2, three models' mathematical expressions and explicit formulas are introduced. After introducing the explicit formulas of the models, the unknown parameters in the formulas are estimated by \sim nls function with R language in Section 3. Then, the final results of the three models and their plots for the Canada population data are found out. Moreover, in Section 3, the mathematical theory pertaining to the function denoted as \sim nls is explained. In Section 4, two different examine methods which are residual plots and AIC are used for three models.

2 Growth models

2.1 Exponential model

The mathematical expression of exponential model, with the value k as the growth rate, P as the population amount with time t is[8]:

$$\frac{\partial P}{\partial t} = k \cdot P$$

Explicit formula of having P in the left hand side:

$$\rightarrow P(t) = C \cdot e^{k \cdot t}$$

2.2 Von Bertalanffy model

The mathematical equation of the Von Bertalanffy model where a , b , γ are the unknown parameters[2]:

$$\frac{\partial P}{\partial t} = a \cdot P^\gamma - b \cdot P$$

The explicit formula of it[2]:

$$P(t) = \left(\frac{a}{b} + (P_0^{1-\gamma} - \frac{a}{b}) \cdot e^{-b(1-\gamma)t}\right)^{\frac{1}{1-\gamma}}$$

P_0 here is the Canada population in 1867.

2.3 Dynamic carrying model

The mathematical expression of the Carrying capacity model, with values K as the upper asymptote and a as the growth rate[2]:

$$\frac{dP}{dt} = a \cdot P \cdot \log\left(\frac{K}{P}\right) \rightarrow$$

Explicit formula:

$$P(t) = C \cdot e^{(a \cdot t \cdot (\log(K) + \log(t) + 1))}$$

C in this formula is a constant which is also an unknown parameter.

3 Estimation of unknown parameters

3.1 Canada population data

Based on the data provided by the Statcan, the constructed table below illustrates the population growth in Canada from 1867 to 2007 with a frequency of 10 years[4]:

Year	Population
1867	3,463,446
1877	4,064,000
1887	4,626,000
1897	5,122,000
1907	6,411,000
1917	8,060,000
1927	9,637,000
1937	11,045,000
1947	12,551,000
1957	16,610,000
1967	20,378,000
1977	23,725,921
1987	26,448,855
1997	29,907,172
2007	32,976,026

The data from this table will be used to find the unknown parameters of the three models. In the real process every single year's population amount will be used instead of ten year base in this table.

3.2 NLS function in R

In this project, the “~nls” function in R coding language will be used in order to derive the parameters for the models. This “~nls” function means “nonlinear least squares” and has been derived mathematically. The nonlinear least squares function is a method commonly used for estimating the parameters of a mathematical model that describes the relationship between a set of independent variables and a dependent variable. Nonlinear least squares are typically used when the relationship between the variables cannot be described by a linear model. This is done by adjusting the parameter values until the sum of the squared errors is minimized. According to the book “Nonlinear regression” written by Seber and Wild[8], first of all, we take some function of y , which represents a nonlinear model

$$y_i = f(x_i; \beta^*) + e_i$$

where β^* is the vector of unknown parameters to be estimated, and $i = 1, 2, 3, \dots, n$, e_i is an error which is normally distributed, and x_i is the vector of predictor variables. Given a set of observed data points (x_i, y_i) for $i = 1, 2, \dots, n$, the goal of NLS is to find the values of the unknown parameters $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ that minimize the sum of squared residuals (SSR), defined as:

$$SSR(\beta) = \sum_{i=1}^n (y_i - f(x_i, \beta))^2$$

To solve for the values of β that minimize $SSR(\beta)$, we can use an iterative algorithm such as the Gauss-Newton algorithm. The Gauss-Newton method is an iterative method used in Nonlinear Least Squares (NLS) estimation to update the parameter estimates. It is based on the Taylor series approximation of the nonlinear function $f(x_i, \beta)$ around the current parameter estimates β_0 . Then:

$$y_i = f(x_i, \beta_0) + \left(\frac{\partial f(x_i, \beta_0)}{\partial \beta_i} \right) (\beta - \beta_i)$$

$$\begin{bmatrix} y_1 - f(x_1, \beta_0) \\ y_2 - f(x_2, \beta_0) \\ \dots \\ y_n - f(x_n, \beta_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial f(x_1, \beta_0)}{\partial \beta_1} & \dots & \frac{\partial f(x_1, \beta_0)}{\partial \beta_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x_n, \beta_0)}{\partial \beta_1} & \dots & \frac{\partial f(x_n, \beta_0)}{\partial \beta_m} \end{bmatrix} \begin{bmatrix} \Delta \beta_1 \\ \dots \\ \Delta \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{bmatrix}$$

It can be expressed as:

$$\Delta Y = A(\beta_0) \Delta \beta + \varepsilon$$

$$\min_{\Delta \beta} \varepsilon_1^2 + \dots + \varepsilon_n^2$$

which can be solved by $\beta_{LS} = [A(\beta_0)^T A(\beta_0)]^{-1} A(\beta_0)^T \Delta Y$

$$\beta_{new} = \beta_0 + \beta_{LS}$$

This process would be repeated

$$\min_{\beta} (\sum_{i=1}^n (y_i - f(x_i, \beta))^2)$$

until the proper β is found out [1].

However, this process requires a considerable amount of effort to derive the values mathematically manually using this function, and it is much easier to apply the "`~nls`" function in R coding language since it has the same properties and gives faster results.

3.3 Results of exponential model

The complete R codes for fitting our models to Canada population data and finding the unknown parameters is in the Appendix. Following is the result of finding the unknown parameters of exponential model with the `~nls` function in R language:

```
Formula: Population ~ a * exp(b * Year)

Parameters:
  Estimate Std. Error t value Pr(>|t|)
a 2.906e-07 1.020e-07 2.848 0.00507 **
b 1.616e-02 1.775e-04 91.062 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

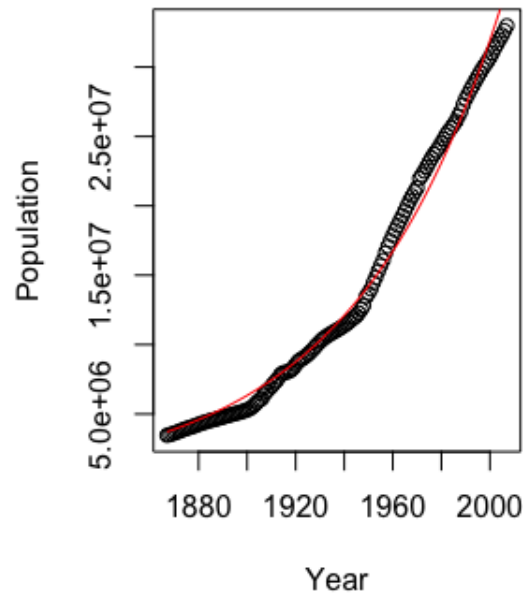
Residual standard error: 964900 on 139 degrees of freedom

Number of iterations to convergence: 103
Achieved convergence tolerance: 1.807e-06
```

The output gave us the values of a and b which are the C and k in our model. Thus, it could be found out that this model fitted to Canada population data is:

$$P = 2.906 \cdot 10^{-7} \cdot e^{1.616 \cdot 10^{-2}}$$

Following is the plot of exponential model which is the red line with the Canada population data:



3.4 Results of Von Bertalanffy model

The output of the Von Bertalanffy model in R language is the following:

Formula: $p \sim (a/b + (p[1]^{(1 - \text{gamma})} - a/b) * \exp(-b * (1 - \text{gamma}) * t))^{1/(1 - \text{gamma})}$

Parameters:

	Estimate	Std. Error	t value	Pr(> t)
a	0.7450	258.7822	0.003	0.998
b	0.6992	258.9927	0.003	0.998
gamma	0.9975	0.8961	1.113	0.268

Residual standard error: 846700 on 138 degrees of freedom

Number of iterations to convergence: 784

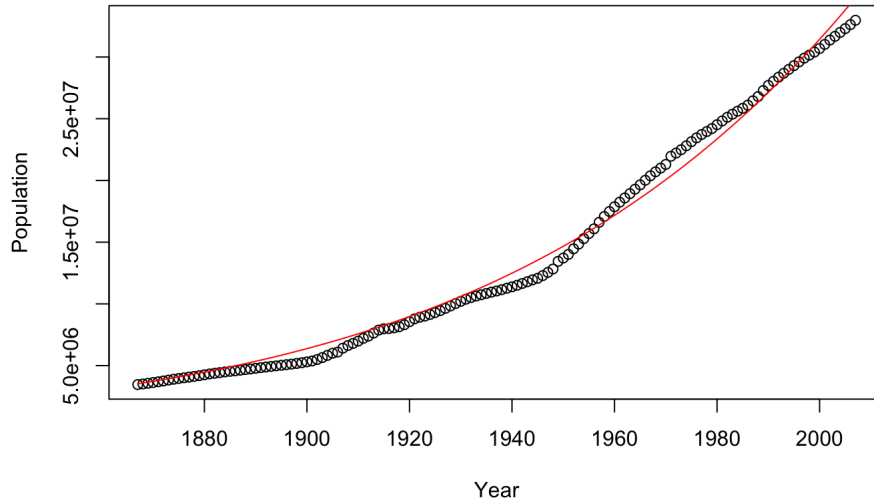
Achieved convergence tolerance: 1.49e-08

The values of a, b and γ are 0.745, 0.6992, and 0.9975 Correspondingly. Therefore, the model of Von Bertalanffy fitted to the data is:

$$P = \left(\frac{0.745}{0.6992} + (P_0^{1-0.9975} - \frac{0.745}{0.6992}) \cdot e^{-0.6992(1-0.9975)t} \right)^{\frac{1}{1-0.9975}}$$

$$\rightarrow P \approx (1.065550343 - 0.0212 \cdot e^{-0.001748t})^{400}$$

Following is plot of the Von Bertalanffy model fitted to the data:



3.5 Results of dynamic carrying capacity model

The output of the dynamic carrying capacity model is:

Formula: $p \sim C * \exp(a * t * (\log(k) - \log(t) + 1))$

Parameters:

	Estimate	Std. Error	t value	Pr(> t)	
a	5.822e-03	6.280e-04	9.270	3.37e-16	***
k	1.427e+03	4.265e+02	3.346	0.00106	**
C	2.267e+06	1.259e+05	18.010	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 752000 on 138 degrees of freedom

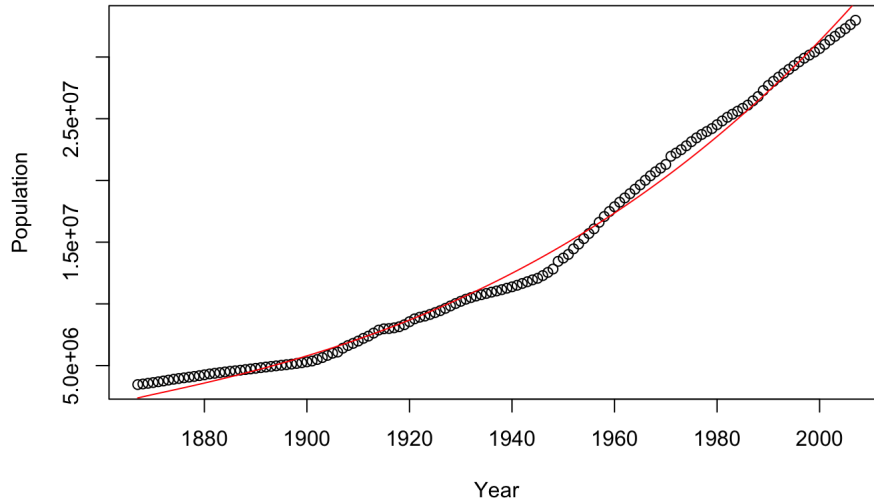
Number of iterations to convergence: 92

Achieved convergence tolerance: 1.49e-08

It found out that the values of the unknown parameters a, k and C. By plugging these values to the model, we got the following formula:

$$P = 2267000 \cdot e^{0.005822 \cdot t(\log(1427) - \log(t) + 1)}$$

The graph of the model with the Canada population data is the following:



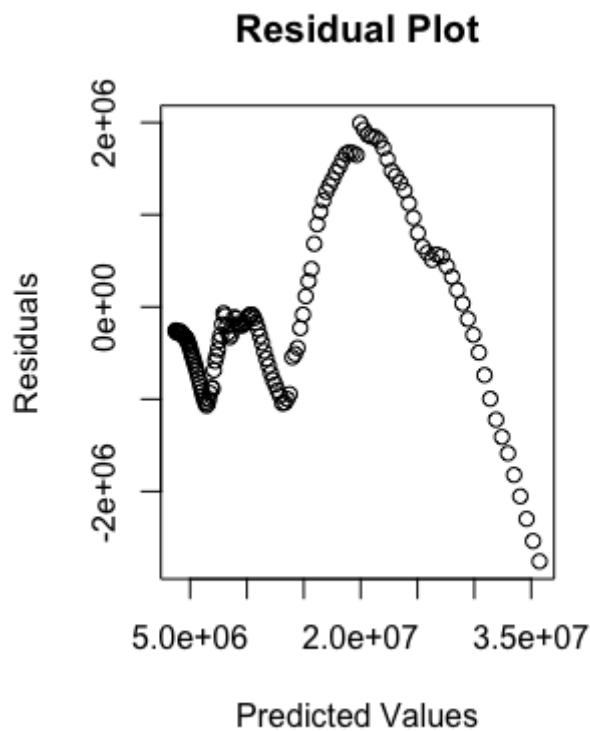
4 Examine methods

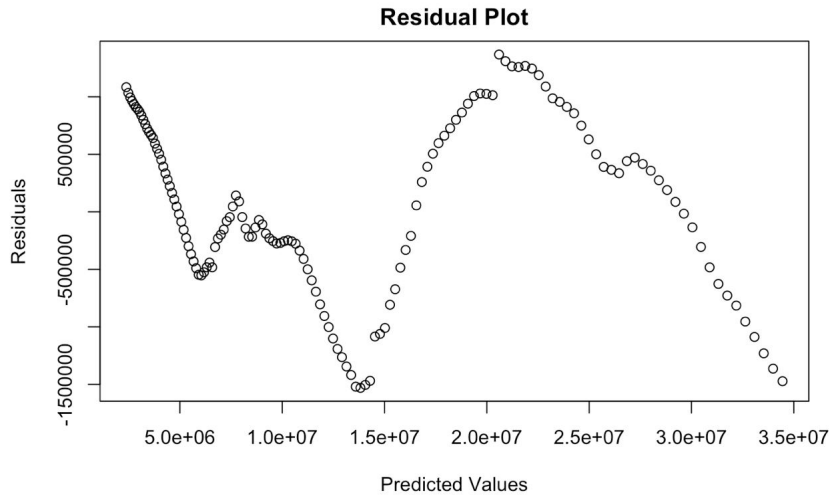
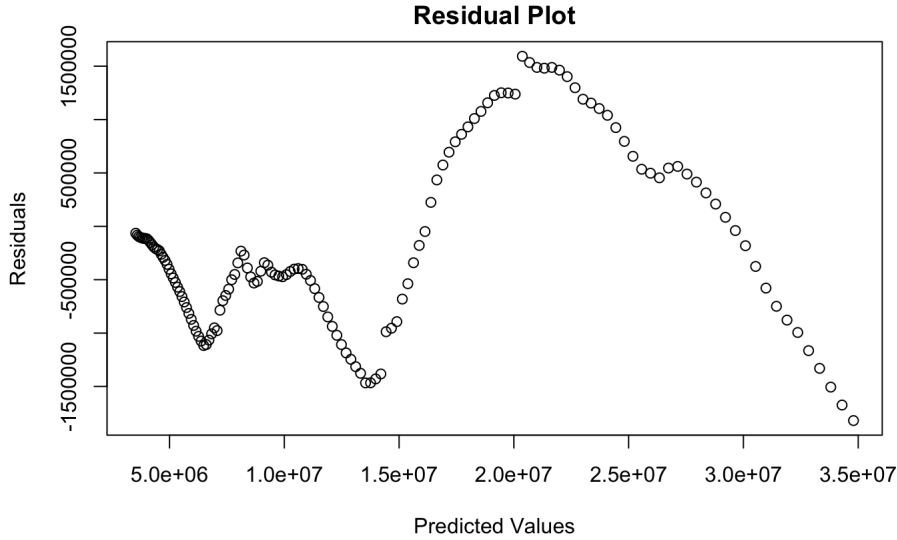
4.1 Residual plots

Application of the residual plot can demonstrate the efficiency of specific modeling. In this paper, it will be used to check the efficiency of Exponential model, Von Bertalanffy model, and dynamic carrying capacity model for the Canada population growth. The formula of residual is:

$$r_i = y_i - \hat{y}_i$$

Where residual r_i are the differences between observed values and the predicted values by the nonlinear models. In case when the residual plot shows some a trend in the graph, then the modeling considered suitable. Because it shows the nonlinear relationship between the predictor variable and the response variable[7]. Following are the residual plots of three models:





First one was the residual plot of Exponential model, second one and third one were Von Bertalanffy model and dynamic carrying capacity model. It is viewed that all three residual plots showed some tendencies which are fluctuating at the beginning and then increasing for a while and decreasing at the end. This is relevant to the growth pattern of the Canada population data. According to the result of Martin et al. in 2017, it could be concluded that these three models are suitable for Canada population data[7].

4.2 AIC method

In order to compare the three models, it is needed to use another compare method which is AIC (Akaike's Information Criteria). AIC (Akaike's Information Criteria) is one of the widespread tools in statistics to estimate whether the model fits the data. Following is how AIC is calculated[3]:

$$AIC = n \log(\hat{\sigma}^2) + 2K$$

Where

$$\hat{\sigma}^2 = \frac{\sum(\hat{\epsilon}_i)^2}{n}$$

$\hat{\epsilon}_i$ here are the residuals estimated in the fitted model, K is the total number of parameters in the model, and n is the recorded measurements[3]. In this paper, to estimate AIC values of three models, R language was used. The R coding for calculating AIC of the three models is also in the Appendix. The AIC value of the exponential model is 4290.033, the Von Bertalanffy model is 4254.152, and the dynamic carrying capacity model is 4220.711. From the results of residual plots, it shows that the three models are all suitable for Canada population data. Moreover, a lower value of AIC indicates a better model fit[3]. Thus, the model with the lowest AIC value is considered to be the best fit. From the results of AIC values, it could be concluded that the dynamic carrying capacity model is the most suitable model for Canada population model with the lowest value of AIC.

5 Future expansion

Nonlinear regression analysis is widely used in population studies to represent the complex linkages between population variables. Due to advances in computing power and statistical methods, future extensions of nonlinear regression analysis have the potential to yield more precise and complex population models. Integrating machine learning algorithms into nonlinear regression models is a potential extension that could lead to more flexible and adaptive models capable of accounting for nonlinear relationships between variables. Nonlinear regression models can incorporate contextual variables such as geographic location and socioeconomic status by using a hierarchical modeling approach. This provides a more complete picture of the factors influencing population dynamics. In addition, advances in Bayesian statistics can account for variability and uncertainty in model parameters, making population models more reliable and resilient. Overall, upcoming modeling improvements could provide scientists with powerful tools to model population dynamics more accurately and better inform policy decisions.

6 Conclusion

This paper has examined the Von Bertalanffy model and the dynamic carrying capacity model based on Canada population data compared to the exponential model. By using the `~nls` function in R language, the unknown parameters in the models were found out. The mathematical theory of how the `~nls` function works was explained. For the purpose of analyzing if the models are proper for the data, two comparison methods were used to test the models which were residual plots and the AIC method. From the results, three models are all fitted well to the data. However, the dynamic carrying model is the most suitable model for the Canada population growth to some extent. In the future, this model could be used as an actuarial life contingency model to calculate different actuarial quantities. This paper showed the suitability of different models to different fields, gave people more possibilities in modeling work.

7 Acknowledgement

I would like to express my heartfelt thanks to all those who supported and inspired me in my capstone project. Firstly, I would like to thank my supervisor, Dr. Wei Dongming, for his guidance, support and encouragement throughout the project. His feedback and advice helped me to define the direction and scope of my research. His support and guidance have been an important part of my academic journey, and I am very grateful for the time and effort he put into my project. Moreover, I want to thank to my second reader Dr. Ugurlu Kerem, for his willingness to review my capstone project as a second reader. His feedback and suggestions gave me a better understanding of the strengths and weaknesses of my work. I would also like to thank the staff at the Nazarbayev university's Mathematics department for providing me with a wealth of knowledge and skills that have greatly contributed to my academic and personal growth. Their passion for mathematics truly inspired me a lot. In addition, I would like to thank my family and friends for their love, support and encouragement throughout my capstone project. Their trust for me has been a source of motivation and inspiration. Thank you all for your contributions to my capstone project.

8 Appendix

R codes for fitting the exponential model to the Canada population, finding the unknown parameters, AIC, and residual plots of the model:

```
setwd("~/downloads")
Data = readxl::read_xlsx("Canada.data.xlsx")
Data = cbind(Time = rownames(Data), Data)
Data$Time = as.numeric(Data$Time)

model1 <- nls(Population ~ a * exp(b * Year), data = Data, start = c(a = 1, b = 0.01), control =
nls.control(maxiter = 104))

plot(Data$Year, Data$Population, pch = 1, xlab = "Year", ylab = "Population")
lines(Data$Year, predict(model1), col = "red")
summary(model1)
residuals <- resid(model1)
plot(fitted(model1), residuals, main = "Residual Plot", xlab = "Predicted Values", ylab = "Residuals")
AIC(model1)
```

R codes for fitting the Von Bertalanffy model and dynamic carrying capacity model to the Canada population, finding the unknown parameters, AIC, and residual plots of the model are the following:

```
library("minpack.lm")
data_used <- read.csv("Canada.csv")
t0 = data_used[,1]
t = data_used[,1]-data_used[1,1]+1
p = data_used[,2]

vb_model <- nlsLM(p ~ (a/b + (p[1]^(1-gamma) - a/b) * exp(-b*(1-gamma)*t))^(1/(1-gamma)),
start = list(a = 2, b=3, gamma=0.01),control=nls.lm.control(maxiter=1024, ftol=1e-6, maxfev=1e6))
plot(t0, p, pch = 1, xlab = "Year", ylab = "Population")
lines(t0, predict(vb_model), col = "red")
summary(vb_model)
AIC(vb_model)
residuals <- resid(vb_model)
plot(fitted(vb_model), residuals, main = "Residual Plot", xlab = "Predicted Values", ylab = "Residuals")

dcc_model <- nlsLM(p ~ C*exp(a*t*(log(k)-log(t)+1)), start = list(a = 5, k=2,C =0.5),control=nls.lm.control(maxiter=1024, ftol=1e-6, maxfev=1e6))

plot(t0, p, pch = 1, xlab = "Year", ylab = "Population")
lines(t0, predict(dcc_model), col = "red")
summary(dcc_model)
AIC(dcc_model)
residuals <- resid(dcc_model)
plot(fitted(dcc_model), residuals, main = "Residual Plot", xlab = "Predicted Values", ylab = "Residuals")
```

9 Reference list

1. Bates, D. M., & Watts, D. G. (2007). *Nonlinear regression analysis and its applications*. New York: John Wiley & Sons.
2. Benzekry, S., Lamont, C., Beheshti, A., Tracz, A., Ebos, J. M., Hlatky, L., & Hahnfeldt, P. (2014). Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Computational Biology*, 10(8). <https://doi.org/10.1371/journal.pcbi.1003800>
3. Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
4. *Estimated population of Canada, 1605 to present* (2015, August 26). Statcan. Retrieved March 30, 2023, from <https://www150.statcan.gc.ca/n1/pub/98-187-x/4151287-eng.htm26>
5. Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4(1), 1-23. <https://doi.org/10.1146/annurev.es.04.110173.000245>
6. Law, M., & Jackson, D. (2014). Residual plots for linear regression models with censored outcome data: A refined method for visualizing residual uncertainty. *MRC Biostatistics Unit*, 3159-3171. <https://doi.org/10.1080/03610918.2015.1076470>
7. Martin, J., Adana, D. D., & Asuero, A. G. (2017). Fitting Models to Data: Residual Analysis, a Primer. *Uncertainty Quantification and Model Calibration*. doi:10.5772/68049
8. Seber, G.A.F., & Wild, C. J. (2003). Nonlinear regression. *John Wiley & Sons, Inc.*, 1-19. <https://doi.org/10.1002/0471725315.fmatter>