Bayesian nonlinear quantile regression using fractional polynomials: COVID-19 UK analysis

Sanna Soomro (Doctoral Researcher) and Keming Yu

Brunel University London, College of Engineering, Design and Physical Sciences

Introduction

The pneumonia was first reported in December 2019 in Wuhan, Hubei Province, China, followed by an outbreak across the country.

As of March 8, 2021, the pandemic of COVID-19 has rapidly spread worldwide with more than 160 countries affected and caused over 117 million infected cases and 2 million deaths.

The UK is the first country to authorise a vaccine against COVID-19. As of January 13, 2021, the UK has authorised three vaccines for COVID-19 that are than any other country in the world. Of which, two vaccines, Oxford and Pfizer vaccines, have been authorised and rolledout across the country whilst the Moderna vaccine has been authorised and is yet to be deployed.

Reproductive number gives a measure of how fast the disease can spread. The higher the reproductive number, the more people will be infected. The R0 of SARS-CoV-2 is estimated to be between 2 and 3.

Methodology

Quantile regression (QR) analysis is a popular regression type, which is robust to extreme outliers.

Rather than using the classical polynomial regression approach, the fractional polynomial (FP) regression model is preferable because it enables the use of root and reciprocal powers allowing the flexibility in model fitting. The constrained set of powers is given by

$$S = \left\{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, 3\right\}.$$

We aim to modelling the nonlinear quantile regression via FP regression that is defined as

$$y_i = \eta(x_i) + \epsilon_i, \qquad i = 1, \ldots, n,$$

where the error term is assumed to follow the asymmetric Laplace (AL) distribution and

$$\eta(x_i) = \beta_0 + \sum_{i=1}^{k} \sum_{j=1}^{m_i} a_{ij} h_{ij}(x_i).$$

Here,

$$h_{ij}(x_i) = \begin{cases} x_i^{p_j}, & if\ p_j \neq p_{j-1}, \\ \log(x_i) h_{i(j-1)}(x_i), & if\ p_j = p_{j-1}, \end{cases}$$

allows the use of repeated powers for the model.

Bayesian formulation: To facilitate the simple Gibbs sampling algorithm, the location-scale mixture representation of the AL can be used to rewrite the error term as

$$\epsilon_i = \theta_1 v_i + \sqrt{\theta_2 v_i \sigma^2} z_i,$$

where

$$v_i \sim Exp(\sigma^{-1}), z_i \sim N(0,1), \theta_1 = \frac{1 - 2\tau}{\tau(1-\tau)}, \theta_2 = \frac{2}{\tau(1-\tau)}.$$

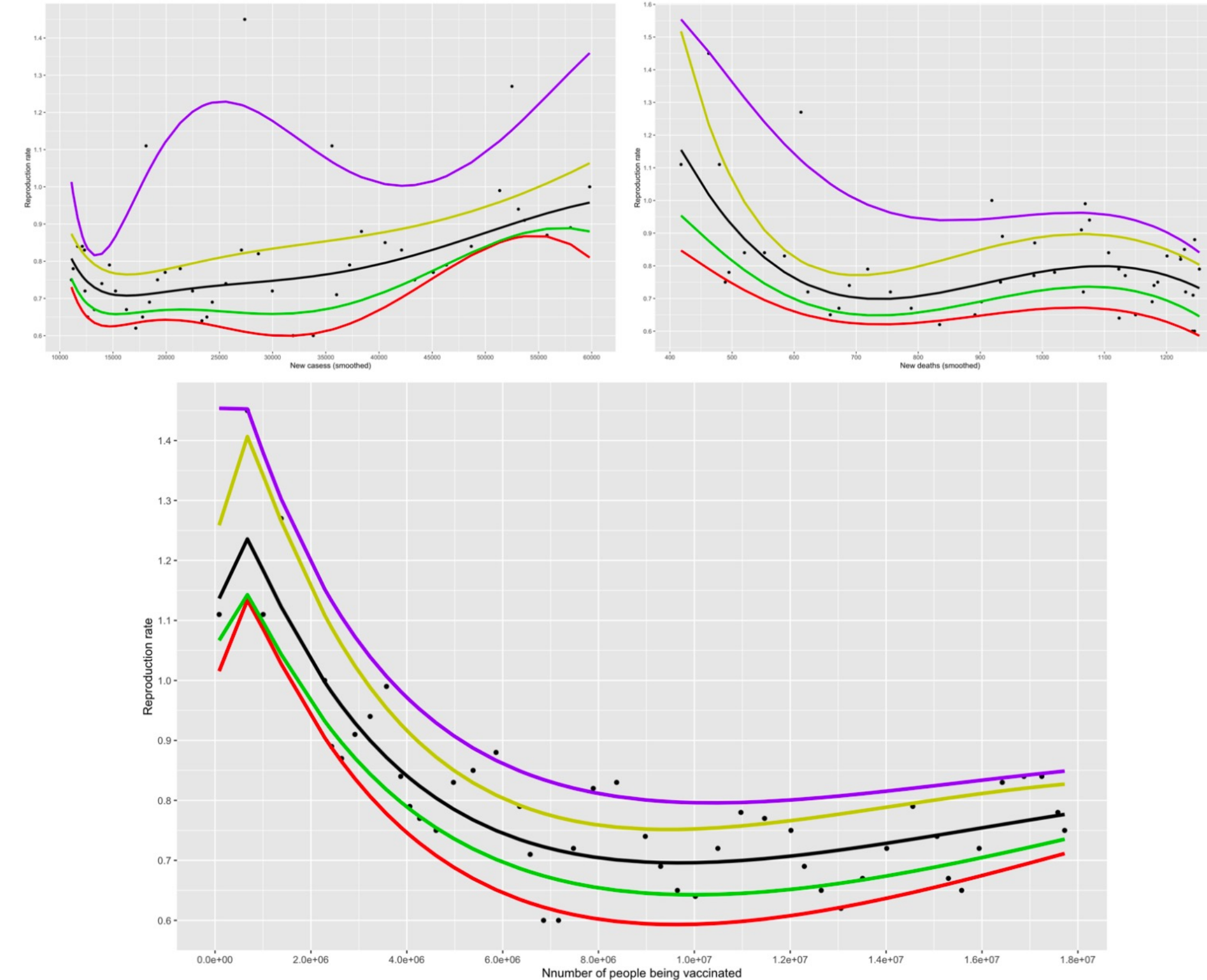## Data Analysis & Results

Figure 1. The fitted lines of Bayesian nonlinear quantile regression under FP model for the univariate case; the first, second and third plots correspond to the fitted model with independent variable *new_cases_smoothed*, *new_deaths_smoothed* and *people_vaccinated*, respectively; the red, green, black, yellow and purple lines correspond to the quantile level of *τ=0.10, 0.25, 0.50, 0.75, 0.90*, respectively

Coefficients	Estimates	Standard Error	t-value	Pr(> |t|)	Significance
Intercept	1.422	8.808×10^{-2}	16.140	$< 2 \times 10^{-16}$	***
new_cases_smoothed	5.756×10^{-7}	1.686×10^{-6}	0.341	0.735	
new_deaths_smoothed	-4.341×10^{-4}	5.260×10^{-5}	-8.252	2.09×10^{-10}	***
people_vaccinated	-2.560×10^{-8}	4.494×10^{-9}	-5.698	1.00×10^{-6}	***
After removing the variable new_cases_smoothed:					
Intercept	1.444	5.863×10^{-2}	24.628	$< 2 \times 10^{-16}$	***
new_deaths_smoothed	-4.277×10^{-4}	4.870×10^{-5}	-8.783	3.12×10^{-1}	***
people_vaccinated	-2.688×10^{-8}	2.493×10^{-9}	-10.778	0.630×10^{-15}	***

Table 1. After carrying out the first LR analysis, the only variable *new_cases_smoothed* is not significant in the model. After implementing another LR analysis excluding this variable, the rest of variables remain remain significant in the model.

		Mean	S.D.	Median	95% Confidence Interval
$\tau = 0.10$	β_0	1.1476	1.0838	0.8573	(-0.1339,2.6782)
	β_1	-0.0005	-0.0006	0.0005	(-0.0013,0.0004)
	β_2	0.0017	0.0141	0.1487	(-0.2660,0.2222)
	β_3 (×10^{-8})	-4.8892	-4.5374	2.1943	(-8.8944,-1.9919)
	β_4	0.0024	0.0020	0.0022	(-0.0005,0.0065)
$\tau = 0.25$	β_0	2.5131	2.5204	0.8693	(1.0878,3.9040)
	β_1	0.0003	0.0003	0.0005	(-0.0005,0.0012)
	β_2	-0.2276	-0.2280	0.1523	(-0.4740,0.0214)
	β_3 (×10^{-8})	-5.4328	-5.1768	2.3502	(-9.5670,-2.0491)
	β_4	0.0031	0.0029	0.0023	(-0.0002,0.0072)
$\tau = 0.50$	β_0	3.8630	3.8524	1.2047	(1.8891,5.8263)
	β_1	0.0009	0.0009	0.0007	(-0.0002,0.0020)
	β_2	-0.4541	-0.4532	0.2128	(-0.8003,-0.1039)
	β_3 (×10^{-8})	-9.2048	-9.4300	3.8613	(-1.5014,-2.5900)
	β_4	0.0065	0.0067	0.0037	(0.0002,0.0120)
$\tau = 0.75$	β_0	4.7890	4.7922	1.0725	(2.9874,6.5324)
	β_1	0.0014	0.0014	0.0006	(0.0003,0.0024)
	β_2	-0.6079	-0.6107	0.1919	(-0.9171,-0.2873)
	β_3 (×10^{-7})	-1.1629	-1.2073	0.3256	(-1.6008,-0.5646)
	β_4	0.0087	0.0092	0.0033	(0.0028,0.0130)
$\tau = 0.90$	β_0	4.8292	4.8170	0.7720	(3.5959,6.1391)
	β_1	0.0014	0.0014	0.0005	(0.0007,0.0022)
	β_2	-0.5966	-0.6001	0.1535	(-0.8482,-0.3393)
	β_3	-1.0282	-1.1057	0.4018	(-1.5323,-2.5945)
	β_4	0.0073	0.0081	0.0041	(-0.0006,0.0124)

Table 2. The estimates of coefficients of Bayesian quantile regression with FP model for the multivariate case.

Data is extracted from the GitHub website for a period from December 13, 2020 to February 21, 2021, and it is managed by Our World in Data. We are interested to examine how the vaccination in the UK can influence the COVID-19 reproduction rate.

Response variable: the COVID-19 *reproduction_rate;*

Covariates: the number of people vaccinated (*people_vaccinated*), the number of new cases (*new_cases_smoothed*), the number of deaths (*new_deaths_smoothed*), all on a daily basis.

Univariate case: The simple FP model is given by
$$y_i = \beta_0 + \beta_1 x_i^{p_1} + \beta_2 x_i^{p_2} + \cdots + \epsilon_i, i = 1, \ldots, n$$

Multivariate case: The multiple FP model is given by
$$y_i = \beta_0 + \beta_1 x_{i1}^{p_{11}} + \beta_2 x_{i1}^{p_{12}} + \cdots + \beta_3 x_{i2}^{p_{21}} + \beta_4 x_{i2}^{p_{22}} + \cdots + \beta_5 x_{i3}^{p_{31}} + \beta_6 x_{i3}^{p_{32}} + \cdots + \epsilon_i, i = 1, \ldots, n,$$

Where $x_1 = new\_cases\_smoothed$, $x_2 = new\_death\_smoothed$ and $x_3 = people\_vaccinated$.

☐ The fitted BQR-FP models clearly exhibit the better relationship between response variable and each independent variable.

☐ The curved lines reflect the nonlinear relationship.

☐ The relationship between *reproduction_rate* and *new_cases_smoothed* sees the expected positive trend.

  ○ τ= 0.10,0.90: the reproduction rate first decreases when the number of daily cases increases by roughly 4000. It then increases and decreases, which reflects a hump shape. After that, it starts to increase again.

  ○ τ= 0.25,0.50,0.75: there is a steady increase within these relationships.

☐ Looking at both *new_deaths_smoothed* and *people_vaccinated*, they have a negative relationship with reproduction rate.

  ○ For some quantiles, there is some increases along the curves.

Discussion

The graph representation for each variable indicates that the use of QR analysis provides a comprehensive picture of the data. The use of FPs allows for the relationships to be nonlinear.

Both the number of people being vaccinated and the number of daily deaths reduce the COVID-19 reproduction rate. This is because, as more people are vaccinated, more people are becoming immunised to COVID-19 and then have a low chance of transmitting or receiving this disease.

Currently, there is no variable selection for Bayesian inference under QR-FP models. This type of method would further improve the model fitting.

References

Bové, D. S. and Held, L. (2011). Bayesian fractional polynomials. *Statistics and Computing* 21, 309–324.

Hideo, K. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of statistical computation and simulation* 81, 1565–1578.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46, 33–50.

Max Roser, Hannah Ritchie, E. O.-O. and Hasell, J. (2020). Coronavirus Pandemic (COVID-19). *Our World in Data* https://ourworldindata.org/coronavirus: accessed on March 11, 2021.

Worldometers.info (2021). Worldometer Dover, Delaware, U.S.A. Worldometers.info: accessed on March 8, 2021.

Zhu, N., Zhang, D., Wang, W., Li, et. al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England journal of medicine* 382, 727–733.