Quantile regression for identifying latent structures in COVID-19 pandemic – Examples from Nepal

Jyoti Devkota[1]

1 Kathmandu University

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

During the COVID-19 pandemic, daily infections exhibited different pattern. It multiplied at an exponential rate, in the beginning. Due to physical restrictions imposed during the lockdown, this number stabilized to a certain value. During the relaxation of lockdowns, the pattern took another form. And after the distribution of three doses of vaccines, this number showed a different trend. In this paper, the path traced by the dependent variable Daily Infected, is explained using quantile values and quantile regression. The time period is from 26 February 2020 to 25 January 2023. Two quantile regression models are developed here. First, quantile regression of daily infection on daily discharged, phase and time of infection and phase time interaction is done for Nepal. Then, quantile regression of daily infection on Ratio 2, phase, time and phase and time interaction is constructed. Ratio 2, is the ratio of total new cases to total deaths, measuring the contribution of total deaths to total infected. The second model is tested for Nepal, India, Germany and the Netherlands. The behavior of the quantiles, before and after vaccination is compared. Here, Germany and the Netherlands are adjoining countries with good quality data. And Nepal and India are taken here as examples of neighboring countries with underreporting of daily infection and deaths. It is found that, quantiles and quantile regression are more robust with respect to underreporting. Thus, the latent behavior of daily incidence of COVID – 19 in different countries with different qualities of data is compared.

Keywords: Under-reporting, Nepal, India, Germany, the Netherlands, pseudo R^2

1. Introduction

In Nepal, the first case of COVID-19 was reported on 23 January 2020. It was a 32-year-old Nepalese student who had recently returned from Hubei, China. The patient recovered, and contacts were also asymptomatic [1]. The Government of Nepal enforced a strict lockdown from 24 March 2020.

COVID-19 pandemic and lock down posed a threat to the healthcare system, in the whole world. In Nepal, government healthcare system was also overwhelmed. It was due to the massive flow of COVID-19 infected into government hospitals. Large number of deaths due to COVID-19 was also a major cause. COVID-19 testing and cure was very expensive in the private sector. So, private healthcare system was beyond the reach of common man [2].

Governmental agencies in Nepal claim that the underreporting in COVID-19 infections and deaths is not very high. They have named the causes to be a) ineffective contact tracing b) lack of coordination in data sharing. Contact tracing was not effective. The number of people testing positive could not be traced. Further, government data sharing from the local level to the central level is very bureaucratic. At that time, this resulted in lack of coordination among various levels of government agencies. The data collection could not be done from a single window. And, there was no harmony between different governmental agencies. Hence in spite of monitoring from World Health Organization (WHO), daily infections and deaths were underreported [3]

In India, more than 4.7 million have died due to COVID-19 between 1 January 2020 to 31 December 2021. This number is around 10 time higher than official records. In the whole world there were 15 million excess deaths during this period [4]. WHO claims that India accounted for almost one third of the COVID-19 deaths globally [5].

Germany and the Netherlands have a good health insurance system. Hence underreporting cases are very low here. Wang et. al have claimed that there is very low underreporting in this region. The ratio between estimated excess mortality rate due to COVID-19 and reported COVID-19 mortality, was between 1 to <2. This was for the cumulative period 2020-21 [6]. During the time of the pandemic, Germany was a standard for sufficient COVID-19 testing[7]. Sufficient COVID-19 testing is related to correct reporting of COVID-19 infections and deaths.

In this paper quantile regression is used to study dynamics of change in COVID-19 pandemic. The spread of the pandemic in the lower and upper quantile values of daily infected is presented. Quantile regression has been used in several data-based research. For example, it has been used in studying the oil price volatility by Liu et al [8]. Different features of impact of crude oil price volatility on shares market price are predicted here. Similarly, S. Majumder et al. used quantile regression in studying the impact of different electricity production sources on CO_2 emission in South Asian countries [9]. N. Das used quantile regression to study influence of industrialization-driven economic transition on carbon footprint in developing nations [10]. Quantile regression was used for predicting the economic sustainability in Vietnam by N. T. Hung [11]. This had to be achieved by realizing opportunities represented by green investment, digitalization, and financial development.

This paper is arranged in following manner. This section is followed by the section Materials and Method, then Result and Discussion and finally Conclusion.

2. Materials and Methods

2.1. Data

For the first model, Nepalese data was taken from the Ministry of Health and Population[12]. For the second model, the data of Germany, the Netherlands, Nepal and India was taken from the WHO [13]. As the data source was the same, results from the four countries could be compared. In both the models, following independent variables are common. They are Phase, Time and interaction between phase and time, represented by Phase: Time. Here, the independent and dependent variables are denoted by capital letters. This is done to facilitate ease of comprehension.

The independent variable Phase takes value from 1 to 35. In the peak of the pandemic, this variable takes the highest values. For Nepal, daily infection was the lowest on 25 January 2023. It was less than 0.005 percent of total infected. Here, Phase was coded as 1. When the daily infected cases were between 0.00959 to 0.01089 percent, then it was coded as 35. It happened when daily infected reached the highest values of 10052. Following the similar scheme, the Phase took values from 1 to 32 for India. For Germany and Netherlands, it took values from 1 to 32 and 1 to 29 respectively.

Similarly, the independent variable Time was classified from 1 to 10. This according to the different control measures imposed by the government. The pre lockdown and strict lockdown period of 2020 was coded as Time 1. Partial lockdown period was coded as Time 2. The complete relaxation of lockdown with certain restrictions was coded as Time 3. Then the second strict lockdown was classified as Time 4. Complete relaxation of second lockdown with certain restrictions was classified as Time 5. The time period of start of Vaccination was coded as Time 6. The start of second wave of COVID – 19 was classified as Time 7. The strict lockdown 3 was classified as Time 8. The relaxation of lockdown 3 with certain restrictions was classified as Time 9. Time 10 was the period when booster dose was distributed in the population. Nepal was taken as a base for all these classifications.

After vaccination period is the period from 1 March 2021 to 25 January 2023. Before vaccination is the period before 1 March 2021. In Nepal vaccination campaign was started on 27 January 2021[14]. The after-vaccination period is same for all the four countries. This is done to facilitate comparison.

2.2. Methodology

Any real valued random variable, X may be characterized by its (right-continuous) distribution function,

$$F(x) = P(X \leq x) \qquad (1)$$

while for any $0 < \tau < 1$,

$$F^{-1}(\tau) = \inf \{x: F(x) \geq \tau\} \qquad (2)$$

τ^{th} quantile of X is represented by Equation (2). Quantile can be defined cut off points, dividing the range of probability distribution into a continuous interval with equal probabilities [15]

$$Q_\tau\left(y_i\right) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \ldots + \beta_p(\tau)x_{ip} \qquad (3)$$

$$i = 1, \ \ldots, \ n$$

Quantile regression is given by Equation (3). Here $Q_\tau(y_i)$ is the τ^{th} quantile of dependent variable Y, where, $0 < \tau < 1$

In quantile regression, the relationship between independent variables and conditional quantile function of dependent variable is quantified. No specific assumption is made regarding the conditional distribution. It hence models the quantiles, instead of the mean as done in standard regression. If there is a violation of assumptions of linear regression like homoscedasticity, use of quantile regression overcomes this limitation. Here interest lies in the outer regions of the conditional distribution. [16]

3. Results and Discussion

It is assumed here that the level of underreporting is constant. This is for every quantile value of dependent variable, Daily Infected. This means that if there exists an underreporting, then it will be same for every quantile. Under this assumption, the path of the Daily Infected modeled by the quantile function, will be unaffected by the underreporting. The path of true values of quantiles will be higher for countries with underreporting. This is in comparison to the path traced by the observed values.

Under this assumption, the path traced by the quantile function of Daily Infected will be same. This will happen irrespective of the level of underreporting. This is also reflected in Figure 1. It is seen that the behavior of quantiles exhibits similar pattern for all the four countries. The behavior of quantiles for period before vaccination and after vaccination is compared in Figure 2. It also exhibits similar pattern.

Also under this assumption, countries with different quality of data can be compared. These quantile regression models can not only explain the incidence of COVID-19 better for countries with underreporting, but also for countries with no underreporting. Further, the interrelationship between the dependent and independent variables represented by the quantile regression coefficients, will be unaffected by this underreporting. Hence, a country wise comparison of the impact different independent variables can be made.

Nepal and India are examples of countries with underreporting of COVID-19 infections and deaths. Nepal and India have adjoining borders. The testing of COVID-19 is expensive for common people of these countries. Governmental health care systems are overburdened. Asymptomatic cases are not detected. Gaps in data transmission between different levels of administration has also resulted reporting error. Hence COVID-19 incidence and death were underreported in these countries.

Germany and the Netherlands are examples of countries with accurate data. These countries are from the developed world and also have adjoining borders. With a compulsory health insurance system, the daily data of COVID-19 incidence and death is accurate for these countries. The reliable and robust system of these countries captures and reports COVID-19 data accurately. Thus, with the use of quantile regression models, countries like India and Nepal (with high underreporting) can be compared with countries like Germany and the Netherlands (with no underreporting).

Daily Infected took a skewed form due to different phases of the pandemic. These phases can be named as the beginning, peak and end of the pandemic. If τ th quantile value of Daily Infected, when τ=0.95 is considered. It means that value of daily infected, for which 95% of the daily infected cases are less than that particular value. So this value lies in the peak of the pandemic. Similarly, when $\tau = 0.01$, this means that only 1% values are lower. This happens at a low point, happening at the end or beginning of the pandemic.

The behavior of Daily Infected is shown by histograms in Figure 3. It is seen that the data is positively skewed for all the four countries. This justifies the significance of quantile regression.

As seen from Figure 3, the frequency distribution of Daily Infected is positively skewed in all the four countries. It is seen that quantile regression can better explain this scenario, especially in the tails. The left tail values on one hand represents drop occurring at the beginning and at the end of the pandemic. The right tail values on the other hand represents the surge at the peak of the pandemic. Period before and after vaccination is also modeled using quantile regression.

The independent variable Ratio 2 of Quantile Regression Model II, is the influence of total deaths on total infections, at a point of time. Ratio 2 takes a finite value as Total Deaths (t) is taken as a non-zero value. This is because the starting date is taken to be the day with non-zero Total Deaths values. This is for the mathematical validity of the variable Ratio 2. Thus, Ratio 2 can be defined in the following manner.

$$\text{Ratio2}(t) = \frac{\textit{Total Infected Cases}(t)}{\textit{Total Deaths}(t)} \qquad (4)$$

1. Ratio 2(t) can be defined as the cumulative effect of deaths denoted by Total Deaths, on Total Infected Cases.
2. If at time t, Total Deaths (t) is very high and Total Infected (t) is also very high, this ratio will be small, this is at the peak of the pandemic and before vaccination.
3. Ratio 2(t) will decrease if the increase Total Deaths is much higher than the increment in Total Infected at t. This happened in the beginning of the pandemic in developed countries like Germany and the Netherlands. These countries have a large elderly population. During this time, a large section of senior citizens of these countries died due to COVID-19. At that time, daily increments in Total Deaths were much higher than daily increment in Total Infected.
4. If at time t, increments in Total Deaths (t) is small and in Total Infected (t) is very large, the increase in this ratio will be large. This happens during the beginning of vaccination period.
5. If at time t, if the increase in Total Deaths and Total Infected is low, the increase this ratio will be small. This happens during and after the vaccination period.
6. Ratio 2 is more robust with respect to underreporting. This is because the underreporting in the numerator cancels out the underreporting in the denominator.

Daily Infected is the dependent variable for Quantile Regression Model I and II. The behavior of Daily Infected and Ratio 2 is exhibited in Table 1. A comparison between the values taken by these variables before and after vaccination is also done. It is seen that for the entire time period, India takes the highest value of average Daily Infected. This is 414188. Nepal has the lowest of 10052. Germany and the Netherlands have values of 307909 and 110432 respectively.

India is followed by Germany, the Netherlands and Nepal. This pattern between the four countries also holds true for the period before vaccination. But for the period after vaccination, Germany has the highest value of Daily Infected, followed by India, the Netherlands and Nepal.

It can also be seen from Table 1, that the average value of Ratio 2 is the highest for the Netherlands and the lowest for India. This is for the entire time period. In the Netherlands, Total Infected is 167.992 times Total Deaths. Whereas in India, it is 71.4034 times Total Deaths. But for the period before vaccination, the average is the highest for Nepal and the lowest for the Netherlands. For after vaccination period, it is the highest for the Netherlands and the lowest for Nepal. Here, Ratio 2 is defined in Equation (4).

Table 1. Behavior of variables Daily Infected Cases and Ratio 2 for four countries

	Nepal	India	Netherlands	Germany	Nepal	India	Germany	Netherlands
Variable	Daily Infected				Ratio 2			
Total								
Mean	1237.687	42596.41	7776.69	35400.15	110.4523	71.40324	90.89984	167.6972
Median	360	16051	2769	12310	83.18126	75.18995	44.06031	106.5933
Standard Deviation	2002.369	73759.81	14435.99	54887.19	77.31882	16.01012	77.73352	142.7757
Minimum	0	3	32	3	69.79242	29.15643	19.21594	7.725747
Maximum	10052	414188	110432	307909	452.1333	92.08087	297	373.3646
Before Vaccination								
Mean	947.6678	31435.29	3019.899	6617.675	183.9021	53.63883	33.89282	32.62387
Median	534	20903	1092	2196	132.5067	58.8172	29.78381	15.64368
Standard Deviation	1028.5	27679.91	3312.373	8135.916	112.4053	15.41202	24.96453	27.91295
Minimum	4	3	32	3	91.16827	29.15643	19.216	7.725747
Maximum	5743	97894	12997	33953	452.1333	81	297	151
After Vaccination								
Mean	1044.451	48257.15	10216.6	50659.83	79.95379	80.41306	121.1234	237.2939
Median	209	13292.5	3346.5	25341	81.90441	82.49318	98.17488	260.4151
Standard Deviation	1978.558	87865.86	17095.05	62482.12	7.313828	4.963176	79.24941	127.346
Minimum	0	79	103	208	69.79242	70.7079	31.5823	69.93314
Maximum	10052	414188	110432	307909	106.2472	92.08087	228.4907	373.3646

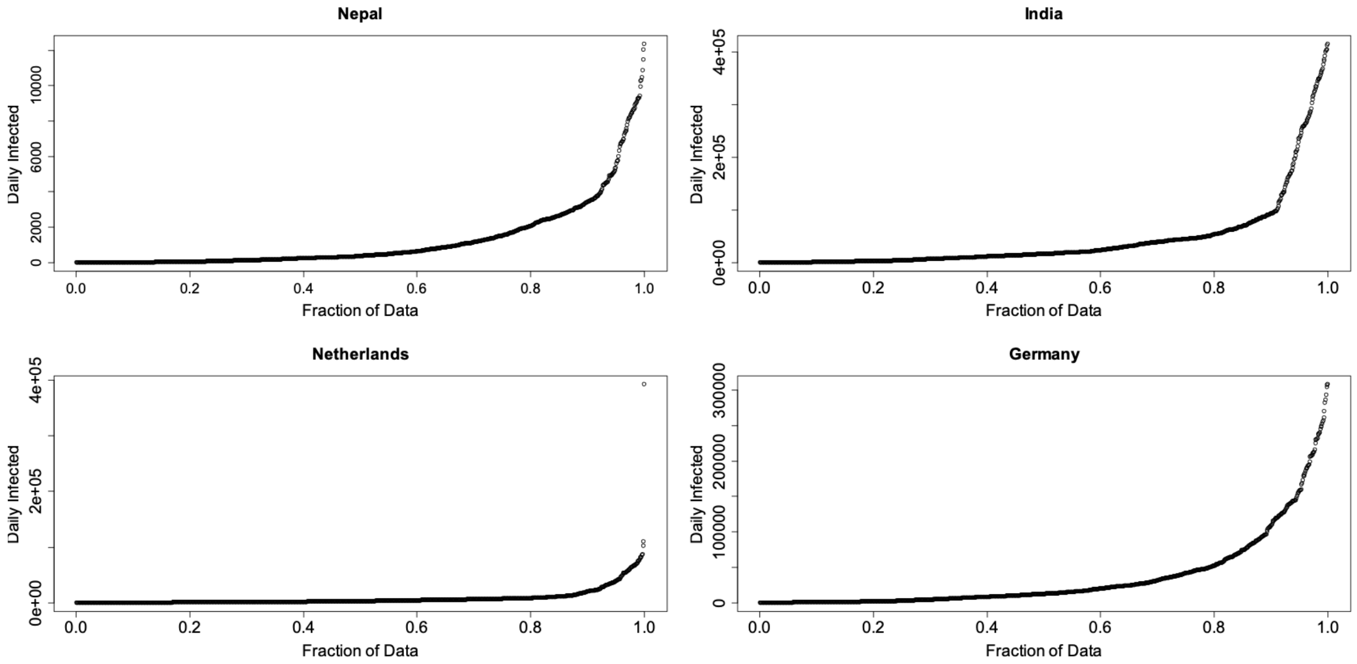

Figure 1. Quantiles of the dependent variable Daily Infected for Nepal, India, Germany and Netherlands

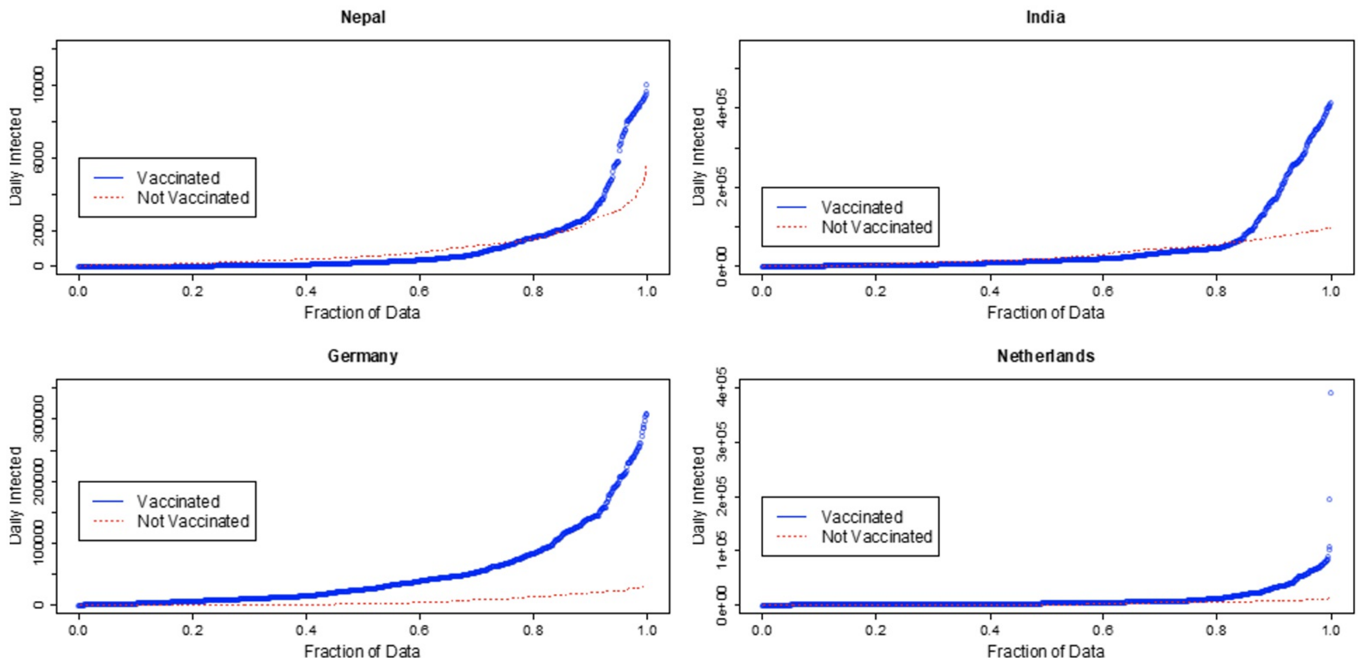

Figure 2. Quantiles of Daily Infected before and after vaccination for Nepal, India, Germany and Netherlands

Here, two quantile regression models are considered. Quantile Regression Model I, explains the behavior of quantile values of daily incidence of COVID-19 for only Nepal. Here the independent variables are Daily Discharged, Phase, Time and interaction between Phase and Time. Quantile regression is also conducted for period before vaccination and period after vaccination, for Nepal. The results are given in Table 2. In Quantile Regression Model II, four countries are considered. They are namely, Nepal, India, the Netherlands and Germany. The independent variables of this model are

Ratio 2, Phase, Time and Phase: Time. The results for this model are given in Table 3.

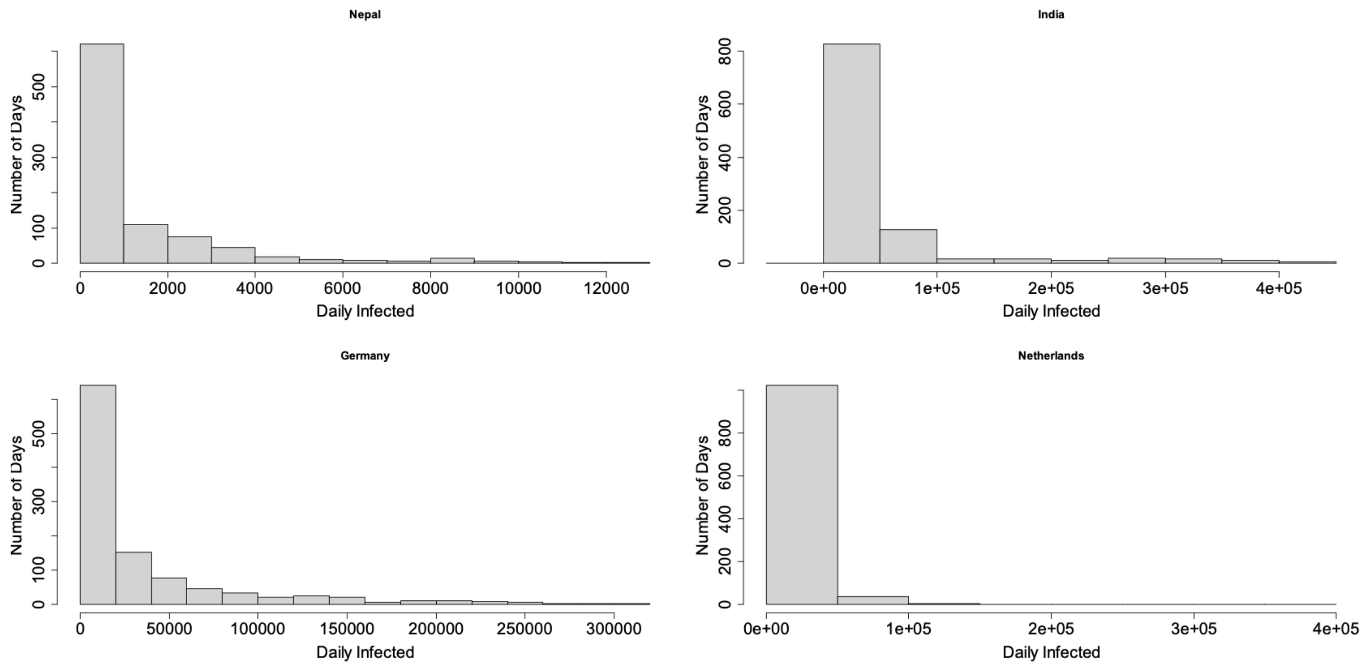

Figure 3. Histogram of dependent variable Daily Infected for Nepal, India, Germany and Netherlands

Daily incidence of COVID-19 is four countries is displayed in Figure 4. The behavior of time series plot of daily incidences of COVID-19 is shown here. The independent variable Day, starts from 25 January 2023, for all the four countries. The arrangement of the independent variable Day is from closest date of 25 January 2023 to farthest date in 2020. For Nepal, it continues till 985, which is 16 May 2020. For India, it is till day 1049 which is 13 March 2020. For Germany, it is till day 1064, which is 26 February 2020. For the Netherlands, it is till day 1055, which is 8 March 2020. Before these dates in 2020, COVID-19 infection had just started. The variable Total Deaths is equal to 1 on 26 February 2020, 8 March, 13 March 2020 and 16 May 2020 for Germany, the Netherlands, India and Nepal respectively. Non zero value of Total Deaths is taken into consideration for the starting date. This is for the mathematical validity of the independent variable Ratio 2.

In Nepal vaccination started on 17 February 2021, for the general public. It was started on a limited scale for front line health workers on 27 January 2021 [17]. Here in this paper, after vaccination period is taken from 1 March 2021 onwards. This date is taken as a base for all the four countries. The vaccination was started on 26 December 2020, 8 January 2021 and 16 January 2021 in Germany, the Netherlands and India, respectively. As we can see, Nepal was the last to start the vaccination drive, among the four countries. So, by the reference date of 1 March 2021, it is assumed that the first dose of COVID-19 vaccination would have some effect in the population. And this effect would be visible in the data collected after this date, in all the four countries. Before vaccination period is till 28 February 2021.

Table 2 and Table 3 give the impact of change in independent variable, on the quantile function of daily infected. This is done by keeping the effects of other independent variables constant. The summary of quantile regression model I, are given in Table 2. The efficiency of quantile models is explained by Goodness of Fit or Pseudo R^2. It is given by Equation (5). This model efficiency parameter takes very high values. It takes values of 0.981, 0.931 and 0.863 for 0.95, 0.5 and 0.05 quantile regressions respectively. It thus explains 98.1, 93.1 and 86.3 percent variability of the data. A linear model fitted to the total data gives and R^2 of 0.997. The pseudo R^2 takes higher values for quantile regression, after vaccination.

Intercept gives the inherent effect of the model. It takes higher values for the period before vaccination and lower values for the period after vaccination. It is highly significant for all the quantiles. Here the intercept of the model is also the estimated conditional quantile function of the daily incidences of COVID-19 with zero values of Discharged Cases/ Ratio 2, Phase, Time and Phase : Time.

It can be seen from Table 2, that the effect of Time is positive for the entire time period. It has a negative effect before vaccination and positive effect after vaccination. The interaction between Phase and Time has a negative effect during the entire time period. It has a positive effect before vaccination and a negative effect after vaccination.

A comparison between the four countries is done in Table 3. Amongst the four countries, Nepal had lowest incidence of daily cases and India had the highest. The Netherlands and Germany were in second and third positions. Accordingly, Intercept takes the highest value for Nepal, followed by the Netherlands, Germany and India. This is for quantile 0.95 and 0.5, which occurs at the peak and the mid of the pandemic. But at the lowest point of the pandemic, represented by quantile 0.05, Intercept can be ordered country wise as the Netherlands, Germany, Nepal and India. This is from the highest to the lowest value.

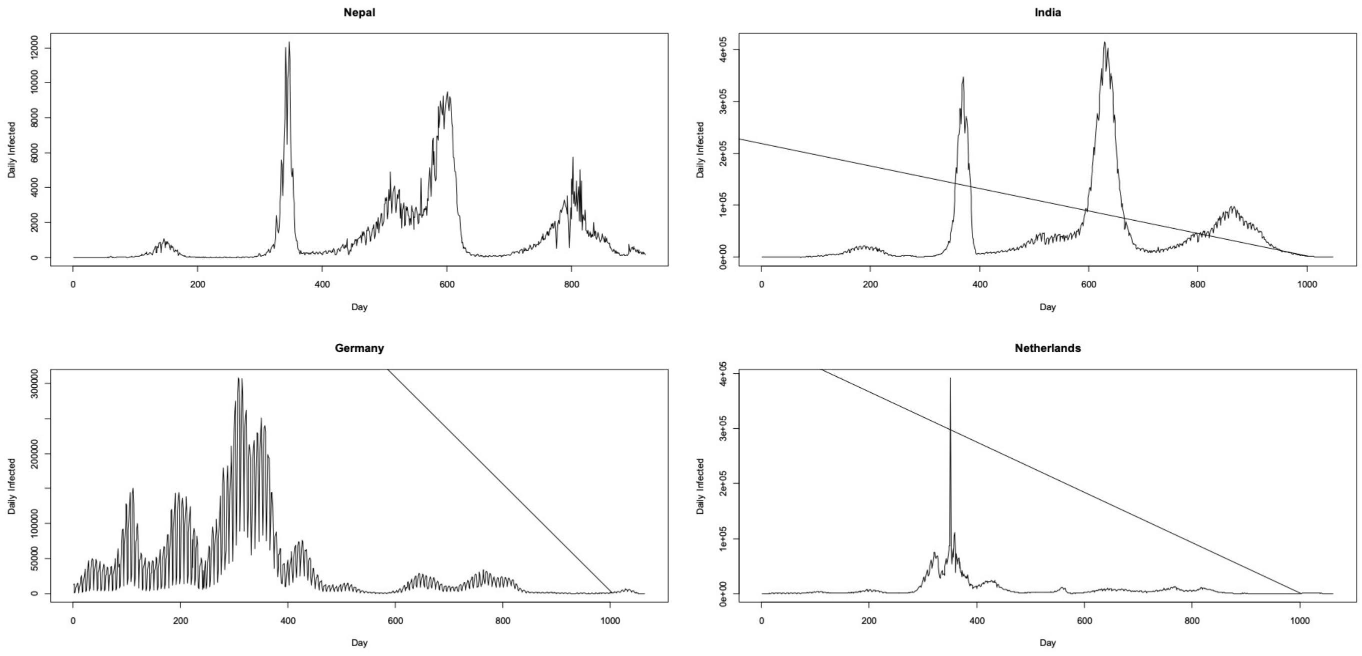

Figure 4. Time series plot of dependent variable Daily Infected for Nepal, India, Germany and Netherlands

$$pseudo\ R^2 = 1 - \left(\frac{\sum_{i=1}^{n} \left| e_i \right|}{\sum_{i=1}^{n} \left| y_i - Q_r \right|} \right) \quad (5)$$

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

$$\hat{y}_i = predicted\ value\ of\ y\ using\ r^{th}\ quantile\ regression$$

$$Q_r = r^{th}\ Quantile\ of\ y$$

Here Y is the dependent variable, daily infected cases.

In other words, pseudo R^2 or goodness of fit for Quantile regression is estimated as 1 minus the ratio of sum of absolute deviations in the fully parametrized models and the sum of absolute deviations in the null (non-conditional) quantile model.

Table 2. Quantile regression model I for Nepal

	$Q_{0.95}$	$Q_{0.5}$	$Q_{0.05}$	**Linear Model of Daily Infected**
Total Period				
Intercept	-348.923***	-585.346***	-613.200***	-0.0520***
Discharged Cases	-0.003	-0.007*	-0.003	-0.001
Phase	304.844 ***	312.550***	303.212***	306.5***
Time	8.675 ***	28.093***	6.953**	15.66***
Interaction between Phase and Time	-0.863 **	-2.096**	-0.622	-1.279***
Goodness of fit for Quantile Regression	0.981	0.931	0.863	0.997 (R^2)
Before Vaccination				
Intercept	-329.736***	-306.814***	-478.904***	-316.4***
Discharged Cases	0.013*	0.0066	-0.004	0.01243
Phase	309.865***	263.027***	267.418***	260.3 ***
Time	-0.127	-33.382*	-19.627	-27.96 **
Interaction between Phase and Time	-2.458	7.531	6.546	7.431 *
Goodness of fit for Quantile Regression	0.954	0.914	0.857	0.9944 (R^2)
After Vaccination				
Intercept	-490.870***	-896.322***	-636.848***	-747***
Discharged Cases	0.00013	-0.009*	0.00072	-0.00109
Phase	310.306 ***	326.360***	312.256***	319.3***
Time	23.334**	60.916***	10.098	39.88***
Interaction between Phase and Time	-1.481**	-3.516**	-1.883	-2.649***
Goodness of fit for Quantile Regression	0.983	0.941	0.864	0.998(R^2)

* Significant at $\alpha = 0.1$, ** Significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$

Table 3. Quantile regression model II

	Q 0.95	Q 0.5	Q 0.05	Linear Model of Daily Infected
Nepal				
Intercept	-363.38439***	-779.59453***	-662.83851***	-629.553**
Ratio2	0.11337	0.58470***	0.14866	0.375**
Phase	303.01493***	323.88460***	304.92733***	311.413**
Time	9.37198	44.85666***	11.96018***	25.700**
Interaction between Phase and Time	-0.76436**	-3.85489***	-0.90600*	-2.027**
Goodness of fit for Quantile Regression	0.9789879	0.925	0.8345142	0.9974 (R^2)
India				
Intercept	-10702.52111***	-19284.04216 ***	-24326.601***	-19052**
Ratio2	-38.53509**	107.03584**	45.858	51.55*
Phase	13004.66586***	12532.68983***	13424.042***	12830.15**
Time	204.80919**	-841.88974***	339.260	-349.85*
Interaction between Phase and Time	43.94634**	80.45666**	-13.378	48.38**
Goodness of fit for Quantile Regression	0.9753632	0.895	0.90600	0.996(R^2)
Germany				
Intercept	-7123.77050***	-4265.55669***	-65.56910	-6434. 979**
Ratio2	-7.54119 ***	-15.12605***	-8.48496***	-8.794**
Phase	9021.67454 ***	5511.84608***	1356.27311*	7061.068**
Time	-261.85367 ***	-937.32079***	-1975.22457***	-875.105**
Interaction between Phase and Time	222.75319 ***	573.31941***	989.89431***	423.779**
Goodness of fit for Quantile Regression	0.9722531***	0.916	0.8536991	0.997(R^2)
The Netherlands				
Intercept	-2080.754***	-1472.713***	-65.346	-2207.532***
Ratio2	3.357**	-1.889***	-0.801***	-1.681
Phase	2527.038***	1684.522***	254.690**	1909.901***
Time	-203.078*	-177.679***	-337.918***	-74.306
Interaction between Phase and Time	68.156	90.624***	189.196***	65.991***
Goodness of fit for Quantile Regression	0.9018	0.833	0.597	0.956(R^2)

* Significant at $\alpha = 0.1$, ** Significant at $\alpha = 0.01$, *** significant at $\alpha = 0.001$

Under the assumption of constant underreporting at each quantile values, the amount of jump in the quantile values between before and after vaccination will be a correct value. As, the errors in these values will cancel out each other. This will irrespective of the level of underreporting. So, this jump will be a correct number for countries with a high level of underreporting. In countries with accurate data, this jump is already a correct number. This jump shows the impact of vaccination on daily infected. This jump is presented in Figure 5 – Figure 9.

The behavior of quantile regression coefficients from quantile 0.1 to 0.9 are given from Figure 5 – Figure 9. Quantile values near to 0.9, depict the peak of the pandemic. And quantile values near 0.1 depict drop in the cases, during the end and beginning of the pandemic. Quantile values near 0.5 depict the median. The progression from waning to surging of daily infections is depicted here. Here, Before and After represents before vaccination and after vaccination respectively.

Following observations are made from the Figure 5 and Figure 6.

1. The path traced by the parameter values before and after vaccination shows a strict demarcation.
2. The inherent effect of this quantile regression model, represented by Intercept, takes a higher value before vaccination and a lower value after vaccination. This means that the impact of COVID-19 has decreased after vaccination.
3. There is a negative impact of the independent variable Time before vaccination and the impact is positive after vaccination, on the dependent variable Daily Infected.
4. The overall impact of Phase is higher after vaccination than before vaccination.
5. The Time and Phase interaction (Time: Phase) has a positive impact before vaccination and negative impact after vaccination.

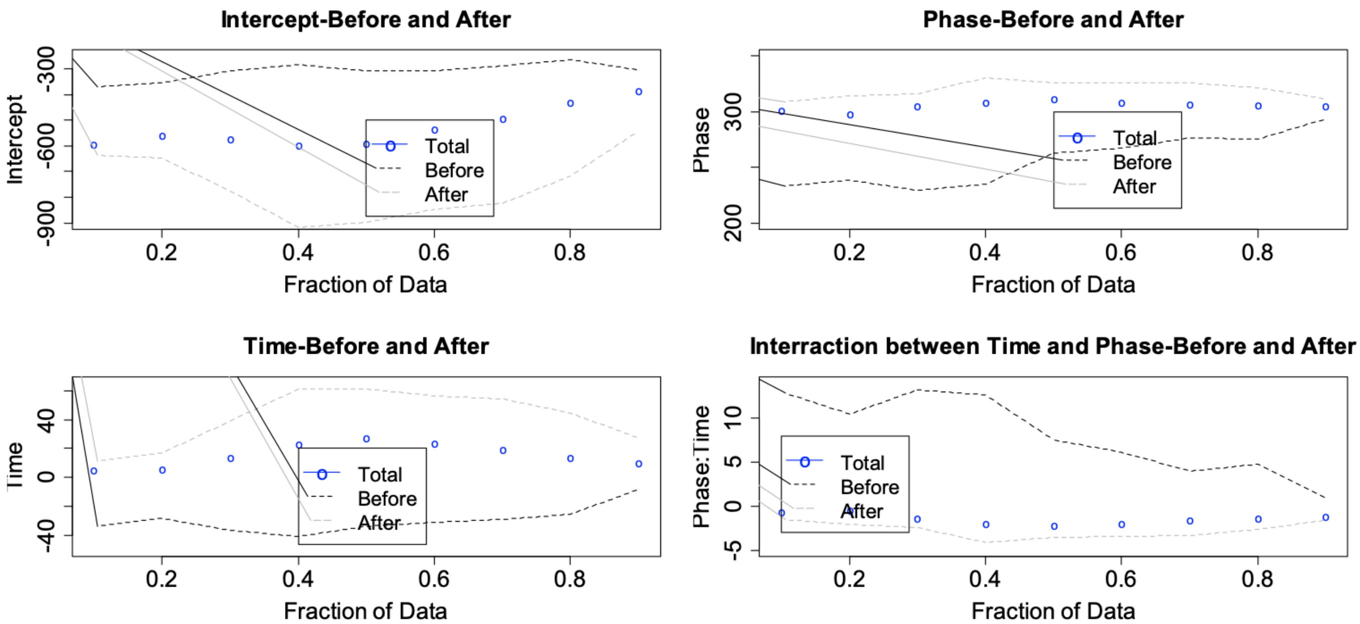

Figure 5. Statistically significant coefficients of quantile regression Model I, Nepal

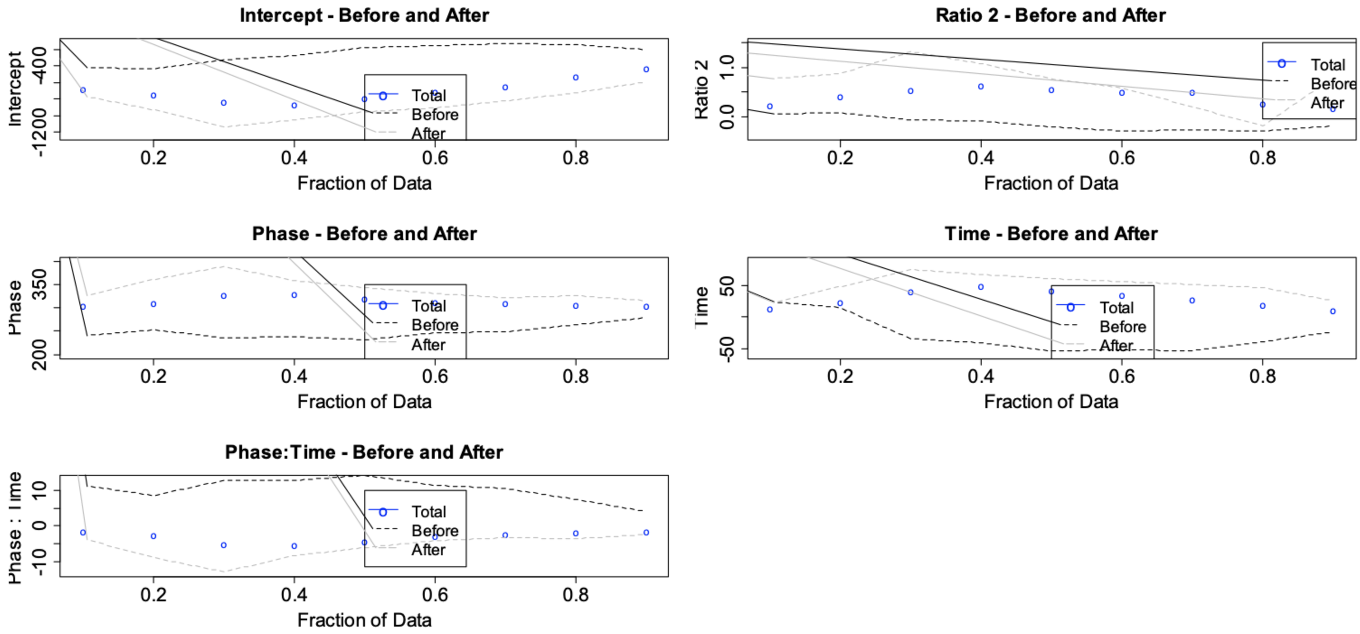

Figure 6. Statistically significant coefficients of quantile regression Model II, Nepal

6. The overall impact of Time and Time and Phase interaction, represented by Total, is positive and negative respectively.
7. It can also be seen that; the impact of Ratio 2 takes higher values after vaccination and lower values before vaccination.

Following interpretations can be made from Figure 7.

1. The values of the intercept drop drastically after vaccination. It takes a higher value before vaccination and a lower value after vaccination. The curves intersect. This is in contrast to Nepal, where the parameter values take separate values before and after vaccination
2. Similarly, the curves traced by regression coefficient of Ratio 2 before and after vaccination also intersect. The impact of Ratio 2, after vaccination takes higher and positive value only after quantile 0.5. Before this quantile, the impact of Ratio 2 is negative.
3. Similar to Nepal, the overall impact of Phase is higher after vaccination than before vaccination.
4. Similarly, like Nepal, Phase: Time takes positive values before vaccination and negative values after vaccination

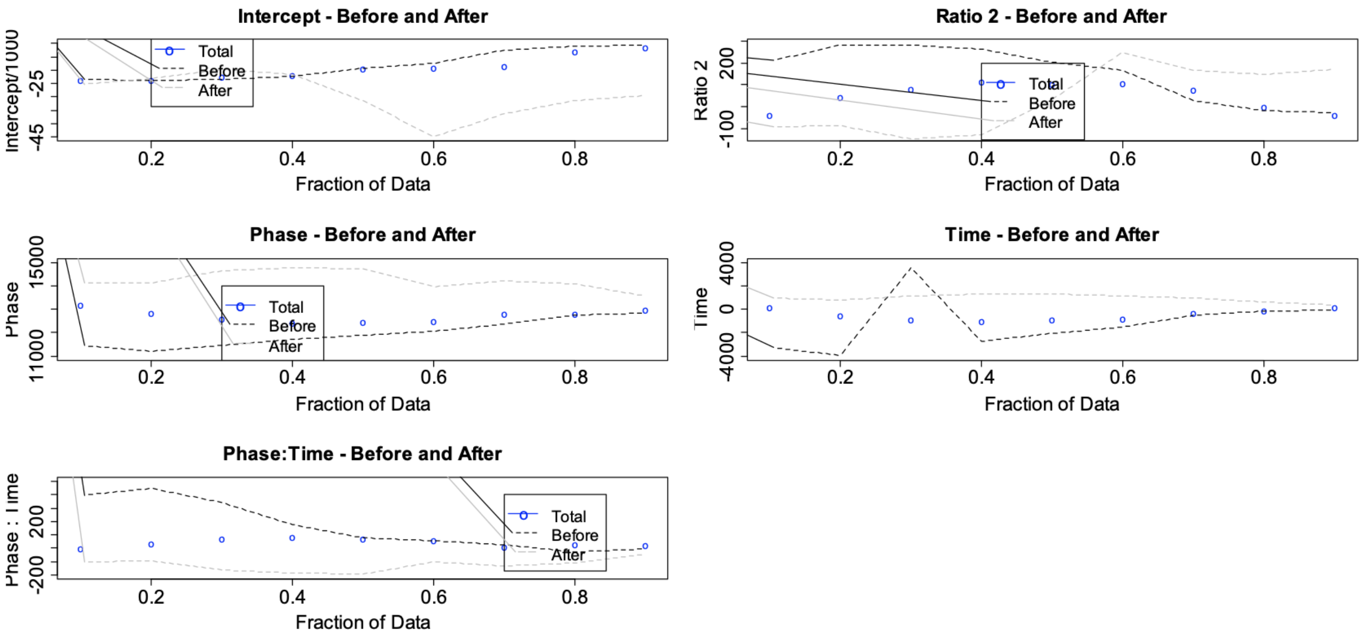

Figure 7. Statistically significant coefficients of quantile regression Model II, India

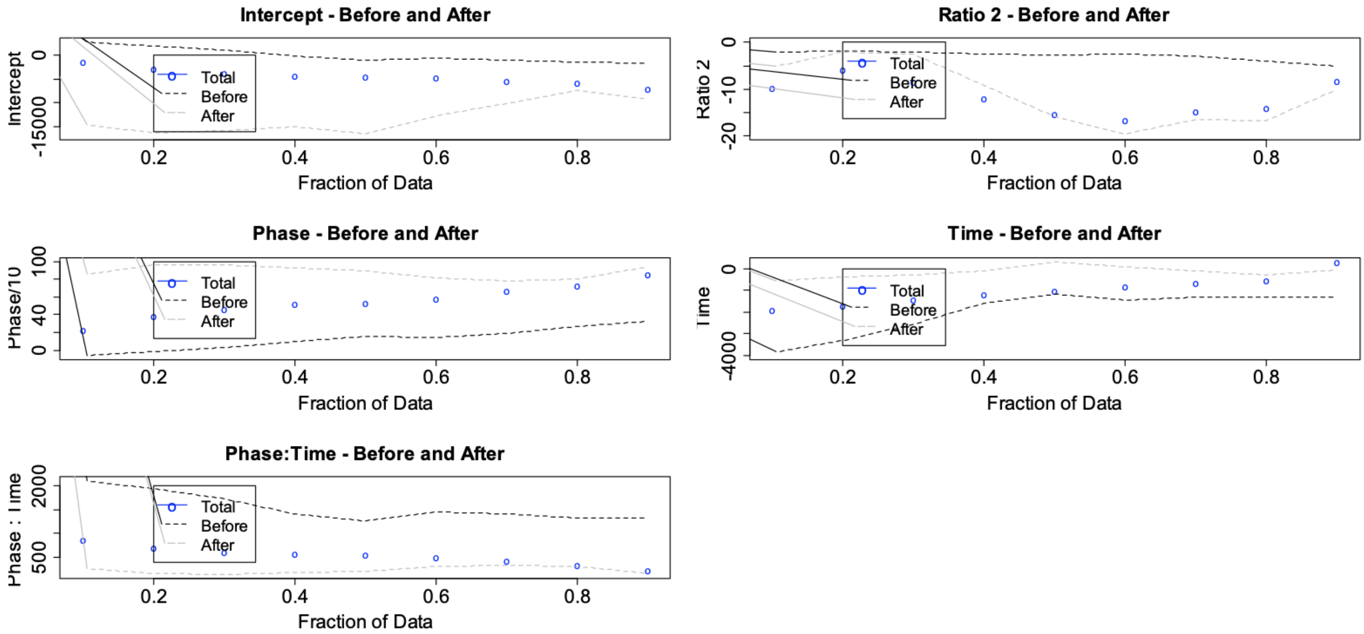

Figure 8. Statistically significant coefficients of quantile regression Model II, Germany

Following interpretations can be made from Figure 8.

1. Like Nepal, the curves traced by Intercept, Ratio 2, Phase, Time and Phase: Time show a strict demarcation. This is for the period before and after vaccination.
2. Like Nepal, the curves traced by the intercept are higher before vaccination than after vaccination.
3. Unlike Nepal, Ratio 2 has a negative impact on the quantile values of Daily Infected. It takes higher values before vaccination than after vaccination.

4. Like Nepal and India, overall impact of Phase is higher after vaccination than before vaccination.
5. Like Nepal, the impact of Time is higher after vaccination than before vaccination.
6. The impact of Phase: Time interaction take positive values before and after vaccination. Like Nepal and India, it takes higher values before vaccination and lower values after vaccination.

Following interpretations can be made from Figure 9.

1. Unlike Nepal and Germany, the curves traced by the parameter values before and after vaccination, intersect. A strict demarcation between these periods doesn't exist.
2. The impact of Phase: Time interaction is higher for period before vaccination than after vaccination.
3. Intercept and impact of Ratio 2, Phase, Time and Phase: Time shows a drastic change for the period after vaccination. This pattern is unique to the Netherlands.
4. The path traced by the coefficients for this country after vaccination, seems to be most sensitive to the quantile values.

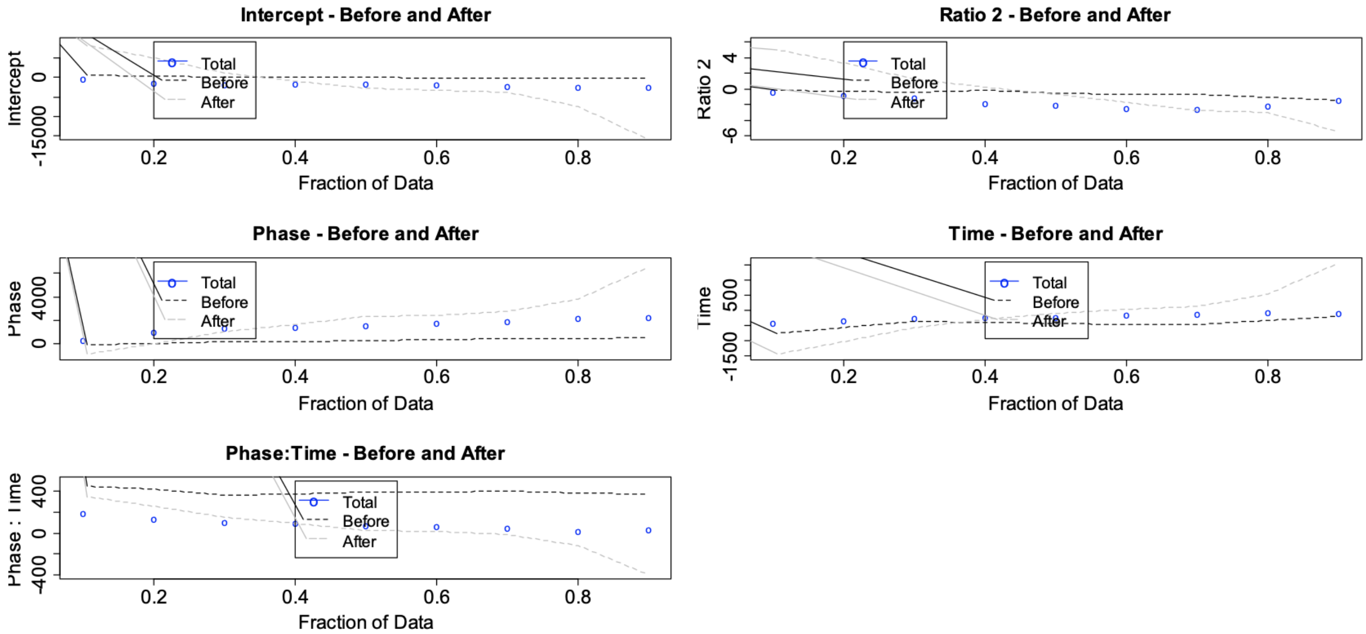

Figure 9. Statistically significant coefficients of quantile regression Model II, the Netherlands

A comparison between the right tail and left tail quantile regression coefficients is done in Figure 10 and Figure 11. This is for Nepal. Right tail values on one hand, show the behavior of the quantiles at the peak of the pandemic. The left tail values on the other hand, show the dip in cases. This is in the beginning and at the end of the pandemic. Quantile Regression model I is considered.

It can be seen from Figure 10, that the intercept surges from the left tail to the right tail. It can also be seen that the impact of Phase remains very high at the right tail than at the left tail. Figure 11 shows that, the impact of the independent variable Time is very high at the right tail than at left tail. It remains very high for all the quantile values at the right tail, except at 0.9958. The effect of this variable at the left tail shows a big fluctuation. The impact of phase and time interraction is very low at the right tail except at the quantile value 0.9958. The influence of the variable Phase and Phase:Time shows a drastic change at the quantile value 0.9958.

The path traced by the regression coefficients of the quantile regression for four countries is displayed in Figure 12 – Figure 14. This for the model II. These figures give a countrywise perspective.

It can be seen from Figure 12, the inherent effect of the model is strictly demarcated at the right tail. Here, the intercept thus takes highest values for Nepal and lowest values for India. This implies that at the peak of the pandemic the paths traced by the inherent effect of the model are separate for these countries. In contrast to this, in the left tail the curves corresponding to Nepal, India and the Netherlands intersect.

Ratio 2 is most sensitive for the Netherlands, in the right tail. This is also shown in Figure 12. In the left tail, the path traced by Ratio 2 is most sensitive for India.

Similarly, we see from Figure 13 that in the right tail, the curve traced by the coefficent of Phase, for different quantiles are non intersecting and stictly demarcated. But in the left tail, the curves for Nepal, Germany and the Netherlands intersect. The impact of time is not strictly demarcated for the four countries, in the left tail as well as right tail. The effect of time on the quantile of daily infected is highest for India, at the left and right tail. As, the path traced by time is the highest for India for almost all quantiles.

Phase and Time interraction one one hand, takes higher value for Germany and the Netherlands, lying in Europe. On the other hand, it takes lower values for Nepal and India. This happens both at the right and the left tail.

5. Conclusion

Direct changes are visible and can be measured. But, inherent and underlying interrelationships also exist. Here these underlying relationships are measured. The focus is on the behavior at the peak and dip of the pandemic. Quantile regression is used to find the conditional quantile function model of Daily Infected, at these points. In order to enable ease of understanding, the independent and dependent variables here, are denoted by capital letters.

It is assumed here that the level of underreporting is same for each quantile value. Under this assumption COVID-19 incidence among four countries is compared. Here Germany and the Netherlands have a good quality data. India and Nepal have incidences of underreporting.

Two models namely Quantile Regression Model I and Quantile Regression Model II are developed. The model accuracy parameters took very high values in case of both the models. It is seen in the Quantile Regression Model I and II, the behavior of quantiles for the period before and after vaccination is strictly demarcated for Nepal.

The quantile regression coefficients for the two models are analysed in detail. It is seen in Quantile Regression Model II that, the intrinsic effect of the model represented by the Intercept, is the highest for Nepal. The second and the third highest values of the Intercept are for the Netherlands and Germany respectively. It takes the lowest value for India. Among these four countries, Nepal has the lowest daily incidence of COVID-19. In contrast to this, India has the highest incidence. This behavior is clearly demarcated in the right tail than in the left tail.

Here, a new variable Ratio 2 is developed for the second model. Ratio 2 measures the contribution of Total Deaths to Total Infected. For Nepal, the values of regression coefficients corresponding to Ratio 2, are positive for lower quantile values. It is at the right tail. Also, for India and Germany, it takes negative values at the right tail. The behavior of regression coefficients of Ratio 2 for the Netherlands, shows a different trend. If the daily increment to the number of Total Deaths is much higher than that of Total Infections, then Ratio 2 decreases. This may be due to large deaths among elderly population. Due to very large daily deaths among this section in comparison to daily infections, Ratio 2 takes a small value. This pattern is exhibited by Ratio 2 in the beginning of the pandemic. This is the left tail. But in the right tail, signifying the top of the pandemic, path traced by Ratio 2 is most erratic.

Similarly the behavior of the independent variables Phase and Time could be interpreted in following manner. The impact of Phase on the spread of the pandemic is strictly demarcated in the right tail. If daily incidences are very high, Phase takes higher value. Phase values are the highest at the peak of the pandemic. The variable Time classified according the different regulatory measures exercised by the government, is the most sensitive to these measures in the right tail, for the Netherlands. It is also the most sensitive to Phase and Time interaction in the right tail.

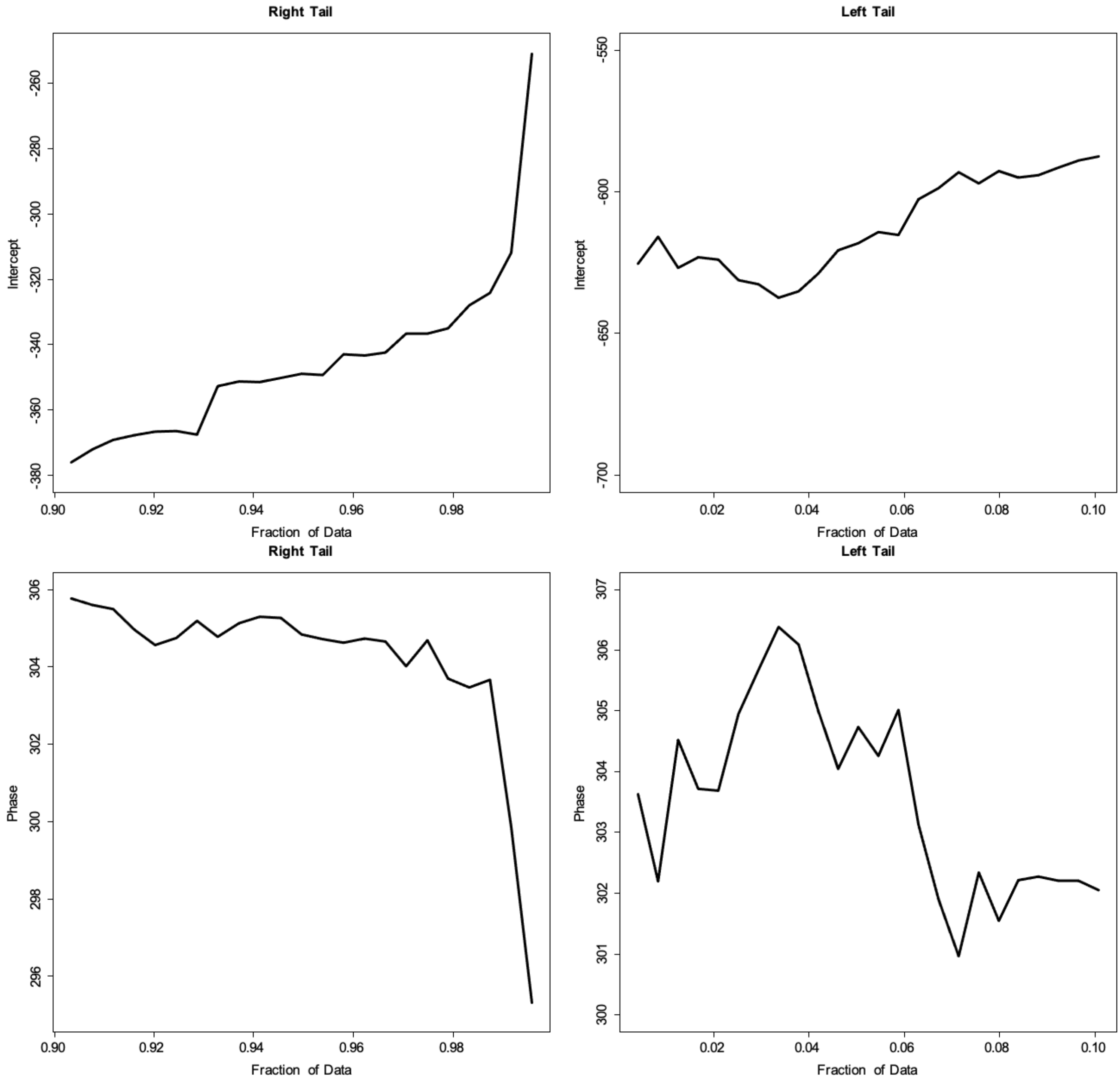

Figure 10. Behavior of Intercept and Phase at the two tails for the Quantile Regression Model I, Nepal

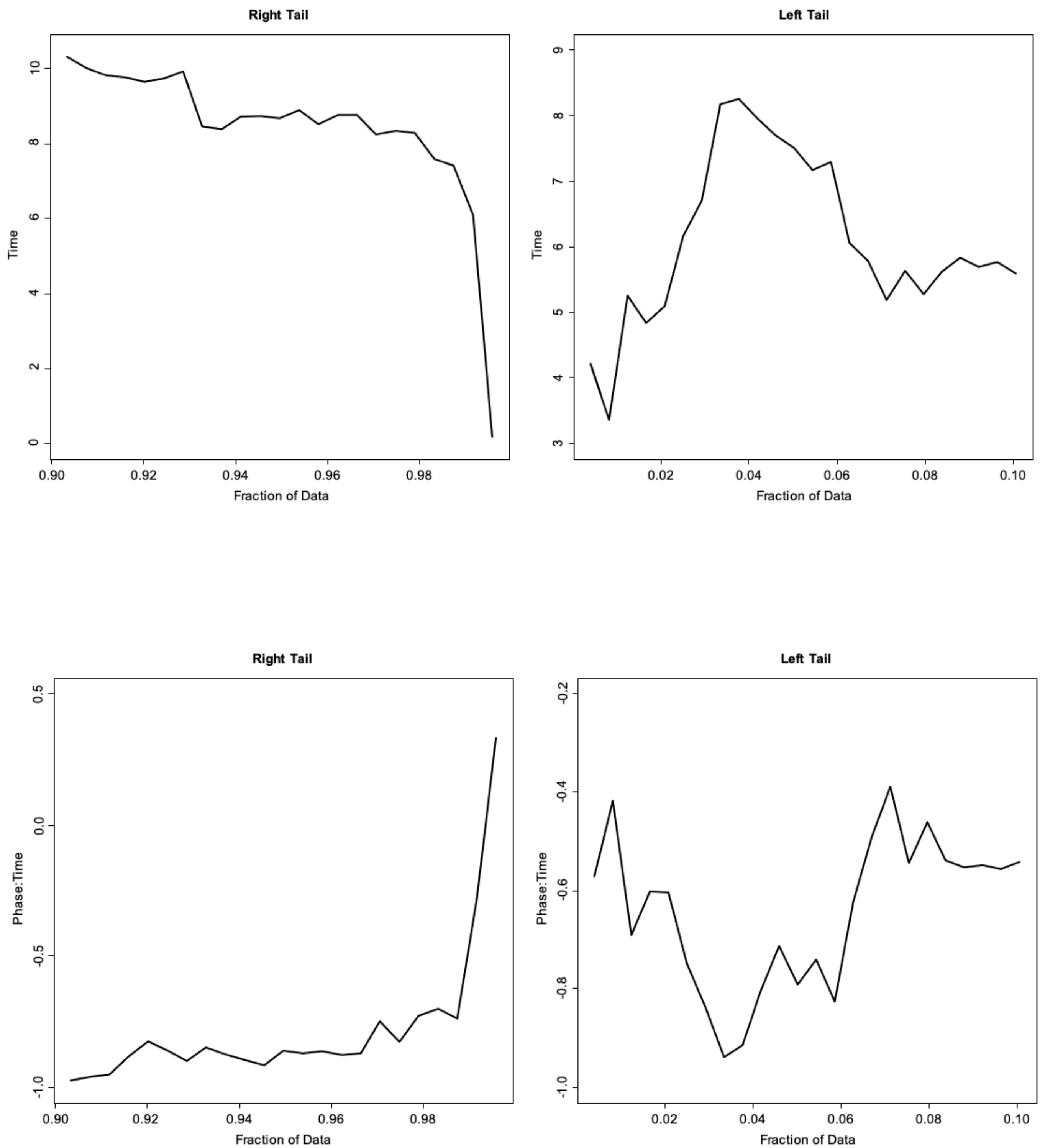

Figure 11. Behavior of Phase and Time interaction at the two tails for the Quantile Regression Model I, Nepal

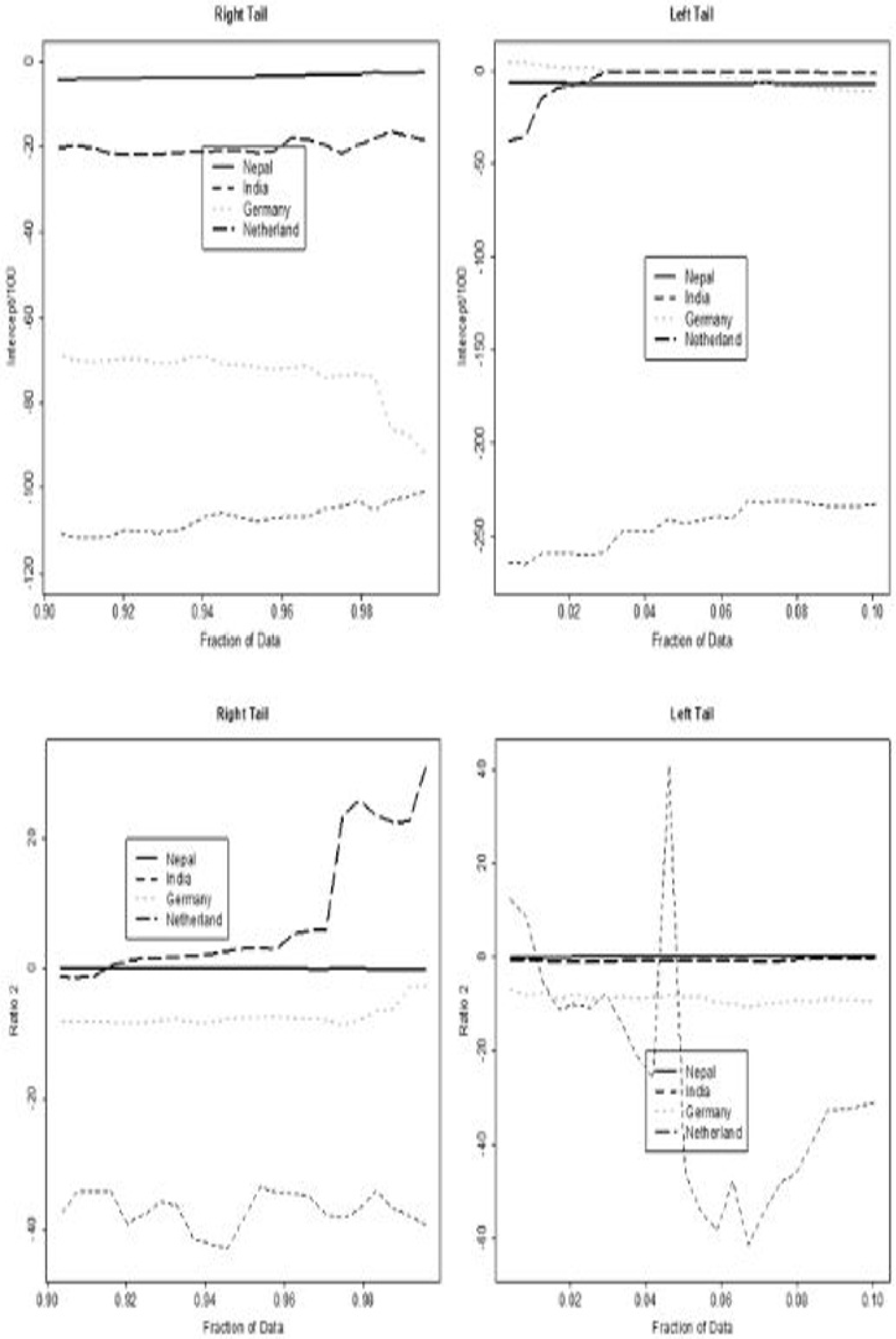

Figure 12. Behavior of Intercept and Ratio 2 at the two tails for the Quantile Regression Model II, country wise

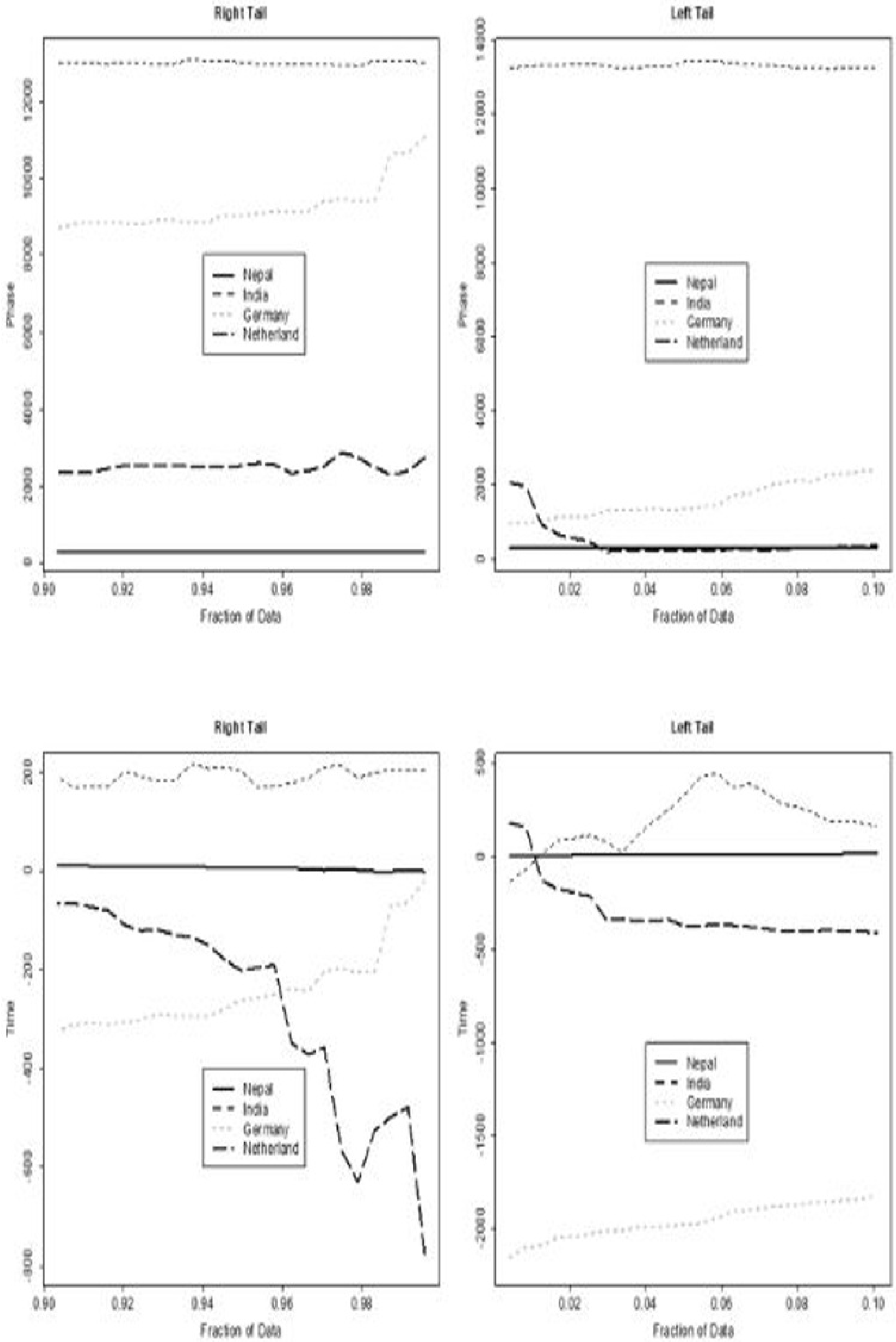

Figure 13. Behavior of Phase and Time at the two tails for the Quantile Regression Model II, country wise

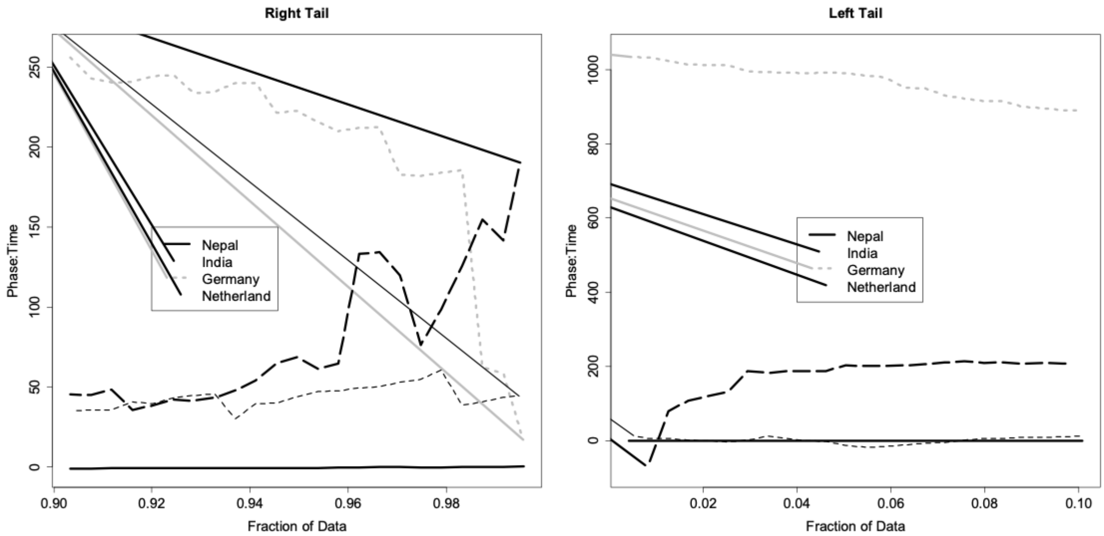

Figure 14. Behavior of Time and Phase: Time at the two tails for the Quantile Regression Model II, country wise

It is seen in the Quantile Regression Model I and II , the period before and after vaccination is strictly demarcated for Nepal.

It is seen in Quantile Regression Model II that, the intrinsic effect of the model represented by the Intercept, is the highest for Nepal. The second and the third highest values of the Intercept are for the Netherlands and Germany respectively. It takes the lowest value for India. Among these four countries, Nepal has the lowest daily incidence of COVID-19. In contrast to this, India has the highest incidence. This behavior is clearly demarcated in the right tail than in the left tail.

Here, Ratio 2 measures the contribution of Total Deaths to Total Infected. For Nepal, the values of regression coefficients corresponding to Ratio 2, are positive for lower quantile values. It is at the right tail. Also, for India and Germany, it takes negative values at the right tail.

The behavior of regression coefficients of Ratio 2 for the Netherlands, shows a different trend. If the daily increment to the number of Total Deaths is much higher than that of Total Infections, then Ratio 2 decreases. This may be due to large deaths among elderly population. Due to very large daily deaths among this section in comparison to daily infections, Ratio 2 takes a small value. This pattern is exhibited by Ratio 2 in the beginning of the pandemic. This is the left tail. But in the right tail, signifying the top of the pandemic, path traced by Ratio 2 is most erratic.

The impact of Phase on the spread of the pandemic is strictly demarcated in the right tail. If daily incidences are very high, Phase takes higher value. Phase values are the highest at the peak of the pandemic.

The variable Time is classified according the different regulatory measures exercised by the government. The Netherlands seems to be the most sensitive to these measures in the right tail. It is also the most sensitive to Phase and Time interaction in the right tail.

References

1. ^A. Bastola, R. Sah, A. J. Rodriguez-Morales et al., The first 2019 novel coronavirus case in Nepal, The Lancet Infectious Diseases, vol. 20, no. 3 (2020), pp. 279-280, 10.1016/S1473-3099(20)30067-0
2. ^D. R. Singh et al., The perils of COVID-19 in Nepal: implications for population health and nutritional status, J Glob Health. 10 (2020), pp1–4., 10.7189/jogh.10.010378
3. ^The Kathmandu Post, Data-discrepancy-amid-a-virus-crisis-could-spell-more-disaster, The Kathmandu Post, Nepal (2021), https://kathmandupost.com/health/2021/05/13/data-discrepancy-amid-a-virus-crisis-could-spell-more-disaster
4. ^BBC, Covid: World's true pandemic death toll nearly 15 million, says WHO, BBC (2022), https://www.bbc.com/news/health-61327778
5. ^BBC, Why India's real Covid toll may never be known, BBC (2022), https://www.bbc.com/news/world-asia-india-60981318
6. ^Wang et al., Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19 related mortality 2020-21, The Lancet, vol. 339, issue 10334 (2022), pp 1513- 1536, 10.1016/S0140-6736(21)02796-3
7. ^Lau et al., Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters, Pulmonology Journal, vol. 27, issue 2 (2021), pp 110-115, 10.1016/j.pulmoe.2020.05.015
8. ^F. Liu, M. Umair and J. Gao, Assessing oil price volatility co-movement with stock market volatility through quantile regression approach, Resources Policy, Vol. 81(2023), 103375, 10.1016/j.resourpol.2023.103375
9. ^S. C. Majumder et al., A Quantile Regression Analysis of the Impact of Electricity Production Sources on CO2 Emission in South Asian Countries, Strategic Planning for Energy and the Environment, vol. 42, issue 2 (2023), 10.13052/spee1048-5236.4223
10. ^N. Das et al., Investigating the nexus between carbonization and industrialization under Kaya's identity: findings from novel multivariate quantile on quantile regression approach, Environmental Science and Pollution Research, vol. 30 (2023), pp 45796 – 45814.
11. ^N. T. Hung, Green investment, financial development, digitalization and economic sustainability in Vietnam: Evidence from a quantile-on-quantile regression and wavelet coherence, Technological Forecasting and Social Change, vol. 186, part B (2023), 122185, 10.1016/j.techfore.2022.122185
12. ^Ministry of Health and Population (MOHP), COVID-19 Dashboard, Ministry of Health and Population, Government of Nepal, https://covid19.mohp.gov.np/
13. ^World Health Organization (WHO), WHO Coronavirus (COVID-19) Dashboard, World Health Organization, https://covid19.who.int/
14. ^World Health Organization (WHO), Nepal story on COVID-19 vaccine deployment: a good start, World Health Organization (2021), https://www.who.int/about/accountability/results/who-results-report-2020-mtr/country-story/2020/nepal-story-on-covid-19-vaccine-deployment-a-good-start
15. ^R. Koenker, Quantile Regression, Econometric Society Monographs, Cambridge University Press.
16. ^E. Waldmann, Quantile regression: A short story on how and why, Statistical Modelling, vol. 18, no. 3-4 (2018), pp 203 – 2018.

17. ^ *The Kathmandu Post, Nepal begins first phase of COVID – 19 vaccination drive, The Kathmandu Post, Nepal (2021), https://kathmandupost.com/health/2021/01/27/nepal-begins-first-phase-of-vaccination-drive-against-covid-19*